Math 330 - Additional Material

Student edition with proofs

Michael Fochler
Department of Mathematics
Binghamton University

This document contains chapter 11 of the Math 330 lecture notes.

Last update: February 17, 2026

Contents

11 Vectors and Vector spaces

11.1 $\mathbb{R}^n$: Euclidean Space

Most if not all of the material of this chapter with the exception of ch.11.2.2 (normed Vector Spaces) on p.522 is familiar to anyone who took a linear algebra course or, in case of two or three dimensional space, to those who took a course in multivariable calculus.

11.1.1 n–Dimensional Vectors

The following definition of an n–dimensional vector is a special case of a vector, something which will be defined as an element of an abstract **"vector space"** . [1]

Definition 11.1 (n–dimensional vectors)**.**

⋆

Let $n \in \mathbb{N}$. An n**–dimensional vector** is a finite, ordered collection $\vec{v} = (x_1, x_2, \dots, x_n)$ of real numbers $x_1, x_2, \dots, x_n$, n is called the **dimension** of the vector $\vec{v}$. □

Thus, $\vec{v}$ is an element of $\mathbb{R}^n$. [2]

Here are some examples of vectors:

Example 11.1 (Two–dimensional vectors)**.** The two–dimensional vector $\vec{v}$ with coordinates $x = -1.5$ and $y = \sqrt{2}$ is written $(-1.5, \sqrt{2})$ and we have $(-1.5, \sqrt{2}) \in \mathbb{R}^2$. Order matters, so this vector is different from $(\sqrt{2}, -1.5) \in \mathbb{R}^2$. □

Example 11.2 (Three–dimensional vectors)**.** $\vec{v_t} = \left(3 - t, 15, \sqrt{5t^2 + \frac{22}{7}}\right) \in \mathbb{R}^3$ with coordinates $x = 3 - t$, $y = 15$ and $z = \sqrt{5t^3 + \frac{22}{7}}$ is an example of a parametrized vector (parametrized by t). Each specific value of t defines an element of $\in \mathbb{R}^3$, e.g., $\vec{v}_{-2} = \left(5, 15, \sqrt{20 + \frac{22}{7}}\right)$. note that

$$F : \mathbb{R} \to \mathbb{R}^3 \qquad t \mapsto F(t) = \vec{v_t}$$

defines a function from $\mathbb{R}$ into $\mathbb{R}^3$ in the sense of definition (**??**) on p.**??**. Each argument s has assigned to it one and only one argument $\vec{v_s} = \left(3 - s, 15, \sqrt{5s^2 + \frac{22}{7}}\right) \in \mathbb{R}^3$.

Or, is it rather that we have three functions

$$\begin{aligned}
&x(\cdot) : \mathbb{R} \to \mathbb{R} \qquad && t \to x(t) = 3 - t, \\
&y(\cdot) : \mathbb{R} \to \mathbb{R} \qquad && t \to y(t) = 15, \\
&z(\cdot) : \mathbb{R} \to \mathbb{R} \qquad && t \to z(t) = \sqrt{5t^2 + \tfrac{22}{7}},
\end{aligned}$$

[1] The definition of an abstract vector space will be given in Definition 11.5 on p.510.

[2] See Definition **??** (Cartesian Product of three or more sets) on p.**??**) concerning $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n\text{times}}$. .

and $t \to \vec{v_t} = (x(t), y(t), z(t))$ is a vector of three real–valued functions $x(\cdot), y(\cdot), z(\cdot)$?

Both points of view are correct and it depends on the specific circumstances how you want to interpret $\vec{v_t}$. □

Example 11.3 (One–dimensional vectors)**.** A one-dimensional vector has a single coordinate.

For example, $\vec{w_1} = (-3) \in \mathbb{R}^1$ with coordinate $x = -3 \in \mathbb{R}$ and $\vec{w_2} = (5.7a) \in \mathbb{R}^1$ with coordinate $x = 5.7a \in \mathbb{R}$ are one–dimensional vectors. $\vec{w_2}$ is not a fixed number but parametrized by $a \in \mathbb{R}$.

Mathematicians do not distinguish between the one–dimensional vector (x) and its coordinate value, the real number x. For brevity, they will simply write $\vec{w_1} = -3$ and $\vec{w_2} = 5.7a$. □

Example 11.4 (Vectors as functions)**.** An n–dimensional vector $\vec{x} = (x_1, x_2, x_3, \cdots, x_n)$ can be interpreted as a real–valued function

$$\begin{aligned} &f_{\vec{x}}(\cdot) : \{1, 2, 3, \cdots, n\} \longrightarrow \mathbb{R} \qquad m \mapsto x_m, \text{ i.e.,} \\ &f_{\vec{x}}(1) = x_1,\ f_{\vec{x}}(2) = x_2,\ \cdots,\ f_{\vec{x}}(n) = x_n, \end{aligned} \tag{11.1}$$

This can be done in reverse. Any real–valued function $f(\cdot) : \{1, 2, 3, \cdots, n\} \longrightarrow \mathbb{R}$ can be associated with the vector $\vec{v}_{f(\cdot)}$ that lists the function values $f(j)$:

$$\vec{v}_{f(\cdot)} := (f(1), f(2), f(3), \cdots, f(n)) \in \mathbb{R}^n. \ \square \tag{11.2}$$

Definition 11.2 (Transposed matrix)**.**

⋆

Let A be a matrix with m rows and n columns. We will write $A = \big((a_{ij})\big)$ to express that a_{ij} denotes the "cell" at the intersection of row i and column j. ($i \in [1, m]_{\mathbb{Z}}$ and $j \in [1, n]_{\mathbb{Z}}$).

$$A = \begin{bmatrix} a_{11}, & a_{12}, & ..., & a_{1n} \\ a_{21}, & a_{22}, & ..., & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}, & a_{m2}, & ..., & a_{mn}(t) \end{bmatrix}.$$

If A is a matrix with m rows and n columns, and if a_{ij} denotes the "cell" at the intersection of row i and column j, then we denote by $A^\top$ the "flipped" matrix which has row i of A as its i–th column, and column j of A as its j–th row.

In other words, if $A = \big((a_{ij})\big)$ and if $A^\top = \big((a^*_{k\ell})\big)$ then $a^*_{ij} = a_{ji}$ for all $i \in [1, m]_{\mathbb{Z}}$ and $j \in [1, n]_{\mathbb{Z}}$. We call $A^\top$ the **transpose** or **transposed matrix** of A. □

$$A^\top = \begin{bmatrix} a_{11}, & a_{21}, & ..., & a_{m1} \\ a_{12}, & a_{22}, & ..., & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n}, & a_{2n}, & ..., & a_{mn}(t) \end{bmatrix}.$$

Remark 11.1.

We usually do not work with matrix multiplication and do not care whether we think of a vector as a column vector or a row vector. For theoretical reasons, most books on linear algebra define vectors $\vec{x} \in \mathbb{R}^n$ as column vectors, as seen on the right.

$$\vec{x} \;=\; \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Since this is very annoying to typeset column vectors, we will often write instead

$$\vec{x} \;=\; (x_1, \dots, x_n)^\top$$

and indicate by means of the transpose symbol $(\cdot)^\top$ that $\vec{x}$ is a column vector. □

11.1.2 Addition and Scalar Multiplication for n–Dimensional Vectors

Definition 11.3 (Addition and scalar multiplication in $\mathbb{R}^n$).

⋆

Given are two n–dimensional vectors
$\vec{x} = (x_1, x_2, \dots, x_n)$ and $\vec{y} = (y_1, y_2, \dots, y_n)$ and a real number α.
We define the **sum** $\vec{x} + \vec{y}$ of $\vec{x}$ and $\vec{y}$ as the vector $\vec{z}$ with the components

$$z_1 \;=\; x_1 + y_1; \quad z_2 \;=\; x_2 + y_2; \quad \dots; \quad z_n \;=\; x_n + y_n; \tag{11.3}$$

We define the **scalar product** $\alpha\vec{x}$ of α and $\vec{x}$ as the vector $\vec{w}$ with the components

$$w_1 \;=\; \alpha x_1; \quad w_2 \;=\; \alpha x_2; \quad \dots; \quad w_n \;=\; \alpha x_n. \tag{11.4}$$ □

Figure 11.1 describes vector addition.

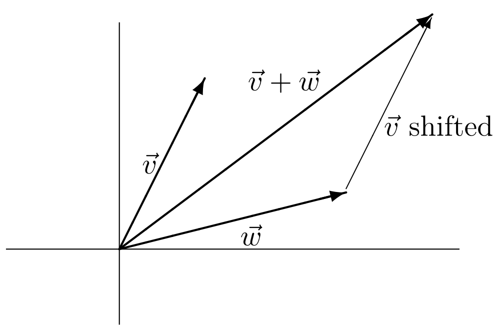


Figure 11.1: Adding two vectors.

Adding two vectors $\vec{v}$ and $\vec{w}$ means that you take one of them, say $\vec{v}$, and shift it in parallel (without rotating it in any way or flipping its direction), so that its starting point moves from the origin to the endpoint of the other vector $\vec{w}$. Look at the picture and you see that the vectors $\vec{v}$, $\vec{w}$ and $\vec{v}$ shifted form three pages of a parallelogram. $\vec{v} + \vec{w}$ is then the diagonal of this parallelogram which starts at the origin and ends at the endpoint of $\vec{v}$ shifted.

11.1.3 Length of n–Dimensional Vectors and the Euclidean Norm

It is customary to write $\|\vec{v}\|_2$ for the length, often also called the **Euclidean norm**, of the vector $\vec{v}$.

Example 11.5 (Length of one–dimensional vectors)**.**
For a vector $\vec{v} = x \in \mathbb{R}$ its length is its absolute value $\|\vec{v}\|_2 = |x|$. This means that $\|-3.57\|_2 = |-3.57| = 3.57$ and $\|\sqrt{2}\|_2 = |\sqrt{2}| \approx 1.414$. □

Example 11.6 (Length of two–dimensional vectors)**.**
We start with an example. Look at $\vec{v} = (4, -3)$. Think of an xy-coordinate system with origin (the spot where x-axis and y-axis intersect) $(0, 0)$. Then $\vec{v}$ is represented by an arrow which starts at the origin and ends at the point with coordinates $x = 4$ and $y = -3$ (see figure 11.2). How long is that arrow?

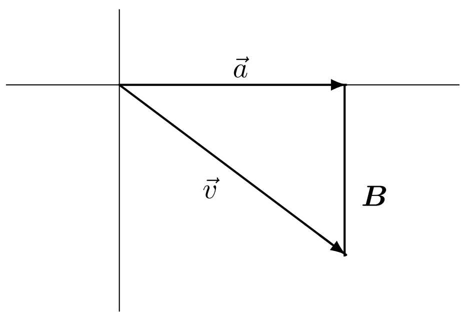


Figure 11.2: Length of a 2–dimensional vectors.

Think of it as the hypothenuse of a right angle triangle whose two other sides are the horizontal arrow from $(0, 0)$ to $(4, 0)$ (the vector $\vec{a} = (4, 0)$) and the vertical line $\boldsymbol{B}$ between $(4, 0)$ and $(4, -3)$. Note that $\boldsymbol{B}$ is not a vector because it does not start at the origin! Obviously (I hope this is obvious) we have $\|\vec{a}\|_2 = 4$ and length–of$(\boldsymbol{B}) = 3$. Pythagoras tells us that

$$\|\vec{v}\|_2^2 \;=\; \|\vec{a}\|_2^2 + \big(\text{length–of–}\boldsymbol{B}\big)^2$$

and we obtain for the vector $(4, -3)$ that $\|\vec{v}\|_2 = \sqrt{16 + 9} = 5$.

The above argument holds for any vector $\vec{v} = (x, y)$ with arbitrary $x, y \in \mathbb{R}$. The horizontal leg on the x-axis is then $\vec{a} = (x, 0)$ with length $|x| = \sqrt{x^2}$ and the vertical leg on the y-axis is a line equal in length to $\vec{b} = (0, y)$ the length of which is $|y| = \sqrt{y^2}$. The theorem of Pythagoras yields $\|(x, y)\|_2^2 = x^2 + y^2$ which becomes, after taking square roots on both sides,

$$\|(x, y)\|_2 = \sqrt{x^2 + y^2} \quad \square \tag{11.5}$$

Example 11.7 (Length of three–dimensional vectors)**.**

This is not so different from the two–dimensional case. We build on the previous example. Let $\vec{v} = (4, -3, 12)$. Think of an xyz-coordinate system with origin (the spot where x-axis, y-axis and z-axis intersect) $(0, 0, 0)$. Then $\vec{v}$ is represented by an arrow which starts at the origin and ends at the point with coordinates $x = 4$, $y = -3$ and $z = 12$. How long is that arrow?

Remember what the standard 3–dimensional coordinate system looks like: The x-axis goes from west to east, the y-axis goes from south to north and the z-axis goes vertically from down below to the sky. Now drop a vertical line $\boldsymbol{B}$ from the point with coordinates $(4, -3, 12)$ to the xy–plane which is "spanned" by the x-axis and y-axis. This line will intersect the xy–plane at the point with coordinates $x = 4$ and $y = -3$ (and $z = 0$. Why?)

Note that $\boldsymbol{B}$ is not a vector because it does not start at the origin! It should be clear that length–of$(\boldsymbol{B}) = |z| = 12$.

We connect the origin $(0, 0, 0)$ with the point $(4, -3, 0)$ in the xy–plane (the endpoint of $\boldsymbol{B}$).

We can apply what we know 2–dimensional vectors because this arrow is contained in the xy–plane. Matter of fact, we have a genuine two–dimensional vector $\vec{a} = (4, -3)$ because the line starts at the origin. Observe that $\vec{a}$ has the same values 4 and -3 for its x– and y–coordinates as the original vector $\vec{v}$. [3] We know from the previous example about two–dimensional vectors that

$$\|\vec{a}\|_2^2 = \|(x, y)\|_2^2 = x^2 + y^2 = 16 + 9 = 25.$$

At this point we have constructed a right angle triangle with **a)** hypothenuse $\vec{v} = (x, y, z)$ where we have $x = 4$, $y = -3$ and $z = 12$, **b)** a vertical leg with length $|z| = 12$ and **c)** a horizontal leg with length $\sqrt{x^2 + y^2} = 5$. Pythagoras tells us that

$$\|\vec{v}\|_2^2 = z^2 + \|(x, y)\|_2^2 \ = \ 144 + 25 = 169, \quad \text{hence} \quad \|\vec{v}\|_2 = 13.$$

None of what we just did depended on the specific values 4, -3 and 12. Any vector $(x, y, z) \in \mathbb{R}^3$ is the hypothenuse of a right triangle where the square lengths of the legs are z^2 and $x^2 + y^2$. We conclude that it is true in general that $\|(x, y, z)\|^2 = x^2 + y^2 + z^2$, hence

$$\|(x, y, z)\| = \sqrt{x^2 + y^2 + z^2} \quad \square \tag{11.6}$$

The previous examples show how to extend the concept of "length" to vector spaces of any finite dimension:

Definition 11.4 (Euclidean norm)**.**

Let $n \in \mathbb{N}$ and $\vec{v} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ be an n–dimension vector. The **Euclidean norm** $\|\vec{v}\|_2$ of $\vec{v}$ is defined as follows:

$$\|\vec{v}\|_2 \quad = \quad \sqrt{{x_1}^2 + {x_2}^2 + \ldots + {x_n}^2} \quad = \quad \sqrt{\sum_{j=1}^{n} {x_j}^2}. \quad \square \tag{11.7}$$

[3] You will learn in the chapter on vector spaces that the vector $\vec{a} = (4, -3)$ is the projection on the xy–coordinates $\pi_{1,2}(\cdot) : \mathbb{R}^3 \to \mathbb{R}^2; \quad (x, y, z) \mapsto (x, y)$ of the vector $\vec{v} = (4, -3, 12)$. See Example 11.20) on p.518.

The above definition is important enough to write the special cases for $n = 1, 2, 3$ where $\|\vec{v}\|_2$ coincides with the length of $\vec{v}$:

$$(11.8)\qquad \begin{aligned} \text{1–dimensional}: &\quad \|(x)\|_2 = \sqrt{x^2} = |x| \\ \text{2–dimensional}: &\quad \|(x,y)\|_2 = \sqrt{x^2+y^2} \\ \text{3–dimensional}: &\quad \|(x,y,z)\|_2 = \sqrt{x^2+y^2+z^2} \end{aligned}$$

Proposition 11.1 (Properties of the Euclidean norm)**.**

Let $n \in \mathbb{N}$. Then the Euclidean norm has the following properties, when viewed as a function

$$\|\cdot\|_2 : \mathbb{R}^n \to \mathbb{R}; \qquad \vec{v} = (x_1, x_2, \dots, x_n) \mapsto \|\vec{v}\|_2 = \sqrt{\sum_{j=1}^{n} {x_j}^2} :$$

(11.9a) $\|\vec{v}\|_2 \geq 0 \quad \forall \vec{v} \in \mathbb{R}^n$ *and* $\|\vec{v}\|_2 = 0 \;\Leftrightarrow\; \vec{v} = 0$ ***(positive definiteness)***

(11.9b) $\|\alpha\vec{v}\|_2 = |\alpha| \cdot \|\vec{v}\|_2 \quad \forall \vec{v} \in \mathbb{R}^n, \forall \alpha \in \mathbb{R}$ ***(absolute homogeneity)***

(11.9c) $\|\vec{v} + \vec{w}\|_2 \leq \|\vec{v}\|_2 + \|\vec{w}\|_2 \quad \forall\, \vec{v}, \vec{w} \in \mathbb{R}^n$ ***(triangle inequality)***

PROOF:

(a) It is certainly true that $\|\vec{v}\|_2 \geq 0$ for any n–dimensional vector $\vec{v}$ because it is defined as $+\sqrt{K}$ where the quantity K is, as a sum of squares, nonnegative. If 0 is the zero vector with coordinates $x_1 = x_2 = \ldots = x_n = 0$ then obviously $\|0\|_2 = \sqrt{0 + \ldots + 0} = 0$. Conversely, let $\vec{v} = (x_1, x_2, \dots, x_n)$ be a vector in $\mathbb{R}^n$ such that $\|\vec{v}\|_2 = 0$. This means that $\sqrt{\sum_{j=1}^{n} {x_j}^2} = 0$ which is only possible if everyone of the nonnegative x_j is zero. In other words, $\vec{v}$ must be the zero vector 0.

(b) Let $\vec{v} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Then

$$\begin{aligned} \|\alpha\vec{v}\|_2 &= \sqrt{\sum_{j=1}^{n} (\alpha x_j)^2} = \sqrt{\sum_{j=1}^{n} \alpha^2 {x_j}^2} = \sqrt{\alpha^2 \sum_{j=1}^{n} {x_j}^2} = \sqrt{\alpha^2}\sqrt{\sum_{j=1}^{n} {x_j}^2} \\ &= \sqrt{\alpha^2}\|\vec{v}\|_2 = |\alpha| \cdot \|\vec{v}\|_2 \end{aligned}$$

because it is true that $\sqrt{\alpha^2} = |\alpha|$ for any real number α (see assumption **??** on p.**??**).

(c) The proof will only be given for $n = 1, 2, 3$.

$\boldsymbol{n = 1}$: (11.9.c) simply is the triangle inequality for real numbers (see (**??**) on **??**) and we are done.

$\boldsymbol{n = 2, 3}$: Look back at the picture about addition of vectors in the plane or in space (see p.505). Remember that for any two vectors $\vec{v}$ and $\vec{w}$ you can always build a triangle whose sides have length $\|\vec{v}\|_2$, $\|\vec{w}\|_2$ and $\|\vec{v} + \vec{w}\|_2$. It is clear that the length of any one side cannot exceed the sum of the lengths of the other two sides, so we get specifically $\|\vec{v} + \vec{w}\|_2 \leq \|\vec{v}\|_2 + \|\vec{w}\|_2$ and we are done.

The geometric argument is not exactly an exact proof but I used it nevertheless because it shows the origin of the term "triangle inequality" for property (11.9.c). An exact proof will be given for arbitrary

$n \in \mathbb{N}$ as a consequence of the so–called Cauchy–Schwartz inequality (cor.11.1). The inequality itself is stated and proved in prop.11.12 on p.524 in the section which discusses inner products (dot products) on vector spaces. ■

11.2 General Vector Spaces

11.2.1 Vector spaces: Definition and Examples

Part of this follows [1] Brin, Matthew and Marchesi, Gerald: Linear Algebra, a text for Math 304, Spring 2016.

Mathematicians are very fond of looking at different objects and figuring out what they have in common. They then create an abstract concept whose items have those properties and examine what they can conclude. For those of you who have had some exposure to object oriented programming: It's like defining a base class, e.g., "mammal", that possesses the core properties of several concrete items such as "horse", "pig", "whale" (sorry – can't require that all mammals have legs). We have looked at the following items that seem to be quite different:

real numbers
n–dimensional vectors
real–valued functions

Well, that was disingenuous. We saw that real numbers and one–dimensional vectors are sort of the same (see 11.3 on p.504). We also saw that n–dimensional vectors can be thought of as real–valued functions with domain $X = \{1, 2, 3, \cdots, n\}$. (see 11.4 on p.504). Never mind, I'll introduce you now to vector spaces as sets of objects which you can "add" and multiply with real numbers according to rules which are guided by those that apply to addition and multiplication of ordinary numbers.

Here is quick reminder on how we add n–dimensional vectors and multiply them with scalars (real numbers) (see (11.1.2) on p.505). Given are two n–dimensional vectors
$\vec{x} = (x_1, x_2, \dots, x_n)$ and $\vec{y} = (y_1, y_2, \dots, y_n)$ and a real number α.
Then the sum $\vec{z} = \vec{x} + \vec{y}$ of $\vec{x}$ and $\vec{y}$ is the vector with the components

$$z_1 = x_1 + y_1; \quad z_2 = x_2 + y_2; \quad \dots \quad z_n = x_n + y_n;$$

and the scalar product $\vec{w} = \alpha\vec{x}$ of α and $\vec{x}$ is the vector with the components

$$w_1 = \alpha x_1; \quad w_2 = \alpha x_2. \quad \dots \quad w_n = \alpha x_n;$$

Example 11.8 (Vector addition and scalar multiplication)**.** We use $n = 2$ in this example:
Let $a = (-3, 1/5)$, $b = (5, \sqrt{2})$ We add those vectors by adding each of the coordinates separately:

$$a + b = (2, 1/5 + \sqrt{2})$$

and we multiply a with a scalar $\lambda \in \mathbb{R}$, e.g. $\lambda = 100$, by multiplying each coordinate with λ:

$$100a = 100(-3, 1/5) = (-300, 20). \ \square$$

In the last example we avoided using the notation "$\vec{x}$" with the cute little arrows on top for vectors. The reason is that this notation is not all that popular in math, even for n–dimensional vectors, and definitely not for abstract vectors as elements of a vector space. Here now is the definition of a vector space, taken almost word for word from the book "Introductory Real Analysis" (Kolmogorov/Fomin [2]). This definition is quite lengthy because a set needs to satisfy many rules to be a vector space.

Definition 11.5 (Vector spaces (linear spaces))**.**

⋆ A nonempty set V is called a **vector space** or **linear space** and we call its elements **vectors** if V satisfies the following:

(A) There exists a binary operation $+ : V \times V \to V;\ (x, y) \mapsto x + y$ on V such that $(V, +)$ is an abelian group (see def. **??** on p.**??**). We call $x + y$ the **sum** of x and y. Note that $(V, +)$ being an abelian group means that the following properties hold for "+":

1. $x + y \ = \ y + x$ for all $x, y \in V$ (**commutativity**);
2. $(x + y) + z \ = \ x + (y + z)$ for all $x, y, z \in V$ (**associativity**);
3. There exists an element $0 \in V$, called the **zero element**, or **zero vector**, or **null vector**, with the property that $x + 0 = x$ for each $x \in V$;
4. For every $x \in V$, there exists an element $-x \in V$, called the **negative** of x, with the property that $x + (-x) = 0$ for each $x \in V$. When adding negatives, then there is a convenient short form. We write $x - y$ as an abbreviation for $x + (-y)$;

(B) There exists a function $\cdot : \mathbb{R} \times V \to V;\ (\alpha, x) \mapsto \alpha \cdot x$, i.e., any real number α and vector x uniquely determine a vector $\alpha \cdot x$. It is customary to simply write αx for $\alpha \cdot x$. This vector is called the **scalar product** of α and x, and it has the following properties:

1. $\alpha(\beta x) \ = \ (\alpha\beta)x$;
2. $1x \ = \ x$;

(C) The operations of addition and scalar multiplication obey the two **distributive laws**

1. $(\alpha + \beta)x \ = \ \alpha x + \beta x$;
2. $\alpha(x + y) \ = \ \alpha x + \alpha y$; □

We state for the reader's convenience the above definition of a vector space V one more time in a more easily remembered form.

Remark 11.2.

★ A vector space V is an algebraic structure with the following properties:

(a) V is nonempty and comes with two assignments:
$+ : V \times V \to V;\ (x, y) \mapsto x + y$, the sum of x and y,
$\cdot : \mathbb{R} \times V \to V;\ (\alpha, x) \mapsto \alpha \cdot x$, (also written αx), the scalar product of α and x.

(c) $(V, +)$ is an abelian group. We write 0 (null vector) for its neutral element, $-x$ for the inverse of a vector x, and $x - y$ for $x + (-y)$.

(d) $\alpha(\beta x) = (\alpha\beta)x$ for all $\alpha, \beta \in \mathbb{R}$ and $x \in V$.

(e) $1 \cdot x = x$ for all $x \in V$. (1 is the real number 1).

(f) Two distributive laws:
$(\alpha + \beta)x = \alpha x + \beta x$,
$\alpha(x + y) = \alpha x + \alpha y$. □

Definition 11.6 (Subspaces of vector spaces)**.**

Let V be a vector space and let $A \subseteq V$ be a nonempty subset of V such that

- For any $x, y \in A$ and $\alpha \in \mathbb{R}$ the sum $x + y$ and the scalar product αx also belong to A.

Then A is called a **subspace** of V.

The set $\{0\}$ which only contains the null vector 0 of V is called the **nullspace**. □

Remark 11.3 (Closure properties of linear subspaces)**.**

(a) Note that if $\alpha = 0$ then $\alpha x = 0$. it follows that the null vector belongs to any subspace.

(b) We ruled out the case $A = \emptyset$ but did not require that A be a strict subset of V ((**??**) on p.**??**). In other words, the entire vector space V is a subspace of itself.

(c) It is trivial to verify that the nullspace $\{0\}$ is a subspace. □

Proposition 11.2 (Subspaces are vector spaces)**.**

A subspace of a vector space is a vector space,
i.e., it satisfies all requirements of definition (11.5).

PROOF: None of the equations that are part of the definition of a vector space magically ceases to be valid just because we look at a subset. The only thing that could go wrong is that some of the expressions might not belong to A anymore. Such can never be the case. Here is the proof for the second distributive law of part C.

We must prove that for any $x, y \in A$ and $\lambda \in \mathbb{R}$

$$\lambda(x + y) = \lambda x + \lambda y.$$

First, $x + y \in A$ because a subspace contains the sum of any two of its elements. It follows that $\lambda(x+y)$ as product of a real number with an element of A again belongs to A because it is a subspace. Hence the left–hand side of the equation belongs to A.

Second, both λx and λy belong to A because each is the scalar product of λ with an element of A and this set is a subspace. It follows for the same reason that the right–hand side of the equation as the sum of two elements of the subspace A belongs to A.

Equality of $\lambda(x + y)$ and $\lambda x + \lambda y$ holds because it holds for x and y as elements of V. ■

Remark 11.4 (Closure properties)**.** If a subset B of a larger set X has the property that certain operations on members of B will always yield elements of B, then we say that B is **closed** with respect to those operations. □

A subspace is a subset of a vector space which is closed with respect to vector addition and scalar multiplication.

You have already encountered the following examples of vector spaces:

Example 11.9 (Vector space $\mathbb{R}$)**.** The real numbers $\mathbb{R}$ are a vector space if you take the ordinary addition of numbers as "+" and the ordinary multiplication of numbers as scalar multiplication. □

Example 11.10 (Vector space $\mathbb{R}^n$)**.** The sets $\mathbb{R}^n$ of n–dimensional vectors become vector spaces if addition and scalar multiplication are defined as in (11.3) on p.505. □

The following example should be thought of as the **definition** of the very important function spaces $\mathscr{F}(X,\mathbb{R})$, $\mathscr{B}(X,\mathbb{R})$, $\mathscr{C}(X,\mathbb{R})$.

Example 11.11 (Vector spaces of real–valued functions)**.** Let X be an arbitrary, nonempty set. Then

$$\mathscr{F}(X,\mathbb{R}) := \{f(\cdot) : f(\cdot) \text{ is a real–valued function on } X\} \tag{11.10}$$

denotes the set of all real–valued functions with domain X [4] and

$$\mathscr{B}(X,\mathbb{R}) := \{g(\cdot) : g(\cdot) \text{ is a bounded real–valued function on } X\}$$

denotes the subset of all bounded real–valued functions with domain X.

Let $A \subseteq \mathbb{R}$. Then

$$\mathscr{C}(A,\mathbb{R}) := \{\psi(\cdot) : \psi(\cdot) \text{ is a continuous real–valued function on } A\}$$

[4] Note that $\mathscr{F}(X,\mathbb{R}) = \mathbb{R}^X$ (see remark **??**, p.**??** which follows Definition **??** of the Cartesian Product of a family of sets.)

denotes the set of all real–valued continuous functions with domain A. [5]

We list separately the case $X = [a, b]$ where $a, b \in \mathbb{R}$ such that $a < b$. Then

$$\mathscr{C}([a,b],\mathbb{R}) \ := \ \{h(\cdot) : h(\cdot) \text{ is a continuous real–valued function for } a \le x \le b\}$$

denotes the set of all contnuous real–valued functions with domain $[a, b]$. Note that, for continuous functions, we had to restrict our choice of domain to subsets of real numbers because there is no notion of continuity for functions on abstract domains (and codomains).

If one defines addition and scalar multiplication as in (**??**) on p.**??**, then each of these sets of real–valued functions becomes a vector space for the following reasons:

I: You can verify properties A, B, C of a vector space by looking at the function values for a specific argument $x \in X$ because then you just deal with ordinary real numbers.

II: The sum of two bounded functions and the product of a bounded function with a scalar is a bounded function. In other words, "$+$" associates with any two elements $f, g \in \mathscr{B}(X, \mathbb{R})$ a third item $f + g \in \mathscr{B}(X, \mathbb{R})$ and "$\cdot$" associates with any $f \in \mathscr{B}(X, \mathbb{R})$ and $\alpha \in \mathbb{R}$ a third item $\alpha \cdot f \in \mathscr{B}(X, \mathbb{R})$.

III: Likewise, the sum of two continuous functions and the product of a continuous function with a scalar is a continuous function. As for bounded functions, "$+$" associates with any two elements $f, g \in \mathscr{C}([a, b], \mathbb{R})$ a third item $f + g \in \mathscr{C}([a, b], \mathbb{R})$ and "$\cdot$" associates with any $f \in \mathscr{C}([a, b], \mathbb{R})$ and $\alpha \in \mathbb{R}$ an item $\alpha \cdot f \in \mathscr{C}([a, b], \mathbb{R})$.

It follows from the above that all three function sets are vector spaces and also that **1)** $\mathscr{B}(X, \mathbb{R})$ is a subspace of $\mathscr{F}(X, \mathbb{R})$, **2)** $\mathscr{C}(X, \mathbb{R})$ is a subspace of $\mathscr{F}(X, \mathbb{R})$.

We will see in ch.**??** (Compactness) on p.**??** that continuous functions defined on a closed interval are bounded. It follows that

$$\mathscr{C}([a,b],\mathbb{R}) \ \subseteq \ \mathscr{B}([a,b],\mathbb{R}) \ \subseteq \ \mathscr{F}([a,b],\mathbb{R}).$$

We deduce from this that **3)** $\mathscr{C}([a, b], \mathbb{R})$ also is a subspace of $\mathscr{B}([a, b], \mathbb{R})$.

It should be noted that, for example, continuous function need **not** be bounded on **open** intervals $]a, b[$, as the example $f(x) = \frac{1}{x}$ demonstrates for $a = 0$ and $b = 1$. □

We summarize the content of the previous example::

Example 11.12 (Vector spaces of real–valued functions)**.**

$$\begin{aligned} \mathscr{F}(X,\mathbb{R}) \ &= \{f(\cdot) : f(\cdot) \text{ is a real–valued function on } X\} \\ \mathscr{B}(X,\mathbb{R}) \ &= \{g(\cdot) : g(\cdot) \text{ is a bounded real–valued function on } X\} \\ \mathscr{C}([a,b],\mathbb{R}) \ &= \{h(\cdot) : h(\cdot) \text{ is a continuous real–valued function for } a \le x \le b\} \end{aligned}$$

- We have subspace relationships $\mathscr{B}(X, \mathbb{R}) \subseteq \mathscr{F}(X, \mathbb{R})$
- We have subspace relationships $\mathscr{C}([a, b], \mathbb{R}) \subseteq \mathscr{B}([a, b], \mathbb{R}) \subseteq \mathscr{F}([a, b], \mathbb{R})$ □

Here are some more examples.

[5] Continuity for such functions was discussed in ch.**??** on p.**??**.

Example 11.13 (Subspace $\{(x,y) : x = y\}$)**.** The set $V := \{(x,x) : x \in \mathbb{R}\}$ of all vectors in the plane with equal x and y coordinates has the following property: For any two vectors $\vec{x} = (a,a)$ and $\vec{y} = (b,b) \in V$ $(a, b \in \mathbb{R})$ and real number α the sum $\vec{x} + \vec{y} = (a+b, a+b)$ and the scalar product $\alpha\vec{x} = (\alpha a, \alpha a)$ have equal x–and y–coordinates, i.e., they again belong to V. It follows that the subset L of $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$ (see (11.6) on p.511). $\square$

A proof for the following is omitted even though it is not difficult:

Example 11.14 (Subspace $\{(x,y) : y = \alpha x\}$)**.** Any subset of the form

$$V_\alpha := \{(x,y) \in \mathbb{R}^2 : y = \alpha x\}$$

is a subspace of $\mathbb{R}^2$ $(\alpha \in \mathbb{R})$. Draw a picture: V_α is the straight line through the origin in the xy–plane with slope α. $\square$

Example 11.15 (Embedding of linear subspaces)**.** The last example was about the subspace of a bigger space. Now we switch to the opposite concept, the **embedding** of a smaller space into a bigger space. We can think of the real numbers $\mathbb{R}$ as a part of the xy–plane $\mathbb{R}^2$ or even 3–dimensional space $\mathbb{R}^3$ by identifying a number a with the two-dimensional vector $(a,0)$ or the three-dimensional vector $(a,0,0)$. Let $m < n$. It is not a big step from here that the most natural way to uniquely associate an n–dimensional vector with an m–dimensional vector $\vec{x} := (x_1, x_2, \dots, x_m)$ by adding zero-coordinates to the right:

$$\vec{x} := (x_1, x_2, \dots, x_m, \underbrace{0, 0, \dots, 0}_{n-m \text{ times}}) \;\square$$

Example 11.16 (All finite–dimensional vectors)**.** Let

$$\mathfrak{S} \;:=\; \bigcup_{n \in \mathbb{N}} \mathbb{R}^n \;=\; \mathbb{R}^1 \cup \mathbb{R}^2 \cup \ldots \cap \mathbb{R}^n \cup \ldots$$

be the set of all vectors of finite (but unspecified) dimension.

We can define addition for any two elements $\vec{x}, \vec{y} \in \mathfrak{S}$ as follows: If $\vec{x}$ and $\vec{y}$ both happen to have the same dimension n then we add them as usual: the sum will be $x_1 + y_1, x_2 + y_2, \dots, x_n + y_n,$. If not, then one of them, say $\vec{x}$ will have dimension m smaller than the dimension n of $\vec{y}$. We define the sum $\vec{x} + \vec{y}$ as the vector

$$\vec{z} := (x_1 + y_1, x_2 + y_2, \dots, x_m + y_m, y_{m+1}, y_{m+2}, \dots, y_n) \;\square$$

Example 11.17 (All sequences of real numbers)**.** Let $\mathbb{R}^{\mathbb{N}} = \prod_{j \in \mathbb{N}} \mathbb{R}$ (see (**??**) on p.**??**). Is this the same set as $\mathfrak{S}$ from the previous example? The answer is No for the following reason: Each element $x \in \mathfrak{S}$ is of some finite dimension, say n, meaning that it has no more than n coordinates. Each element $y \in \mathbb{R}^{\mathbb{N}}$ is a collection of numbers $y_1, y_2, \dots$ none of which need to be zero. In other words, $\mathbb{R}^{\mathbb{N}}$ is the vector space of all sequences of real numbers. Addition is of course done coordinate by coordinate and scalar multiplication with $\alpha \in \mathbb{R}$ is done by multiplying each coordinate with α.

There is again a natural way to embed $\mathfrak{S}$ into $\mathbb{R}^{\mathbb{N}}$ as follows: We transform an n–dimensional vector $(a_1, a_2, \dots, a_n)$ into an element of $\mathbb{R}^{\mathbb{N}}$ (a sequence $(a_j)_{j \in \mathbb{N}}$) by setting $a_j = 0$ for $j > n$. $\square$

Definition 11.7 (linear combinations)**.**

⋆

Let V be a vector space and let $x_1, x_2, x_3, \dots, x_n \in V$ be a finite number of vectors in V. Let $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n \in \mathbb{R}$. We call the finite sum

$$\sum_{j=0}^{n} \alpha_j x_j = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \ldots + \alpha_n x_n \tag{11.11}$$

a **linear combination** of the vectors x_j . The multipliers $\alpha_1, \alpha_2, \dots$ are called **scalars**. □

In other words, linear combinations are sums of scalar multiples of vectors. The expression in (11.11) always is an element of V, no matter how big $n \in \mathbb{N}$ was chosen:

Proposition 11.3 (Vector spaces are closed w.r.t. linear combinations)**.**

Let V be a vector space and let $x_1, x_2, x_3, \dots, x_n \in V$ be a finite number of vectors in V. Let $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n \in \mathbb{R}$. Then the linear combination $\sum\limits_{j=0}^{n} \alpha_j x_j$ also belongs to V. Note that this is also true for subspaces, because those are vector spaces, too.

PROOF: Trivial. ■

Proposition 11.4.

Let V be a vector space and let $(W_i)_{i \in I}$ be a family of subspaces of V. Let $W := \bigcap \big[W_i : i \in I \big]$. Then W is a subspace of V.

PROOF: It suffices to show that W is not empty and that any linear combination of items in W belongs to W. As $0 \in W_i$ for each $i \in I$, it follows that $0 \in W$, hence $W \neq \emptyset$.

Let $x_1, x_2, \dots x_k \in W$ and $\alpha_1, \alpha_2, \dots \alpha_k \in \mathbb{R} (k \in \mathbb{N})$. Let $x := \sum_{j=1}^{k} \alpha_j x_j$. Then $x \in W_i$ for all i because each W_i is a vector space, hence $x \in W$. ■

Definition 11.8 (Linear span)**.**

⋆

Let V be a vector space and $A \subseteq V$. Then the set

$$\mathit{span}(A) \ := \ \{ \ \sum_{j=1}^{k} \alpha_j x_j : k \in \mathbb{N}, \alpha_j \in \mathbb{R}, x_j \in A \ (1 \leq j \leq k) \ \}. \tag{11.12}$$

of all linear combinations of vectors in A is called the **span** or **linear span** of A. □

Proposition 11.5.

Let V be a vector space and $A \subseteq V$. Then span(A) is a subspace of V.

PROOF: Let $y_j \in$ span$(A)$for $j = 1, 2, \dots k$, i.e. y_j is a linear combination of vectors $x_{j,1}, x_{j,2}, \dots x_{j,n_j} \in A$. But then any linear combination of $y_1, y_2, \dots y_k$ is a linear combination of the vectors

$$x_{1,1},\ x_{1,2}, \dots,\ x_{1,n_1},\ x_{2,1},\ x_{2,2}, \dots,\ x_{2,n_2}, \dots,\ x_{k,1},\ x_{k,2}, \dots,\ x_{k,n_k}. \blacksquare$$

Theorem 11.1.

Let V be a vector space and $A \subseteq V$.

Let $\mathscr{V} := \{W \subseteq V : W \supseteq A$ and W is a subspace of $V\}$. Then span$(A) = \bigcap [W : W \in \mathscr{V}]$.

PROOF: Clearly, span$(A) \supseteq A$ It follows from prop.11.5 that span$(A) \in \mathfrak{V}$, hence span$(A) \supseteq \bigcap [W : W \in \mathfrak{V}]$.

On the other hand, Any subspace W of V that contains A also contains all its linear combinations, hence span$(A) \subseteq W$ for all $W \in \mathfrak{V}$. But then span$(A) \subseteq \bigcap [W : W \in \mathfrak{V}]$. $\blacksquare$

Remark 11.5 (Linear span(A) = subspace generated by A)**.** Let V be a vector space and $A \subseteq V$. Theorem 11.1 justifies to call span(A) the **subspace generated by** A. ☐

Definition 11.9 (linear mappings)**.**

⋆

Let V_1, V_2 be two vector spaces. Let the function $f(\cdot) : V_1 \to V_2$ satisfy

(11.13a)	$f(x+y) = f(x) + f(y) \quad \forall x, y \in V_1$	**additivity**
(11.13b)	$f(\alpha x) = \alpha f(x) \quad \forall x \in V_1,\ \forall \alpha \in \mathbb{R}$	**homogeneity**

Then we call $f(\cdot)$ a **linear function** or **linear mapping**. ☐

Note 11.1 (Note on homogeneity)**.** We encountered "absolute homogeneity" when examining the properties of the Euclidean norm ((11.9) on p.508). That is not the same concept as homogeneity for linear functions because you had to take the absolute value $|\alpha|$ instead of α. ☐

Remark 11.6 (Linear mappings are compatible with linear combinations). We saw in the last proposition that vector spaces are closed with respect to linear combinations. Linear functions and linear combinations work harmoniously in the following sense:

(A): The image of the sum is the sum of the images,

(B): The image of the scalar multiple is the scalar multiple of the image,

(C): The image of the linear combination is the linear combination of the images.

In other words, linear mappings preserve or are structure compatible with linear combinations. Matter of fact, **(A)** asserts that f is a homomorphism $(V_1,+) \to (V_2,+)$ from the group $(V_1,+)$ to the group $(V_2,+)$. See Definition **??** (Homomorphisms and isomorphisms) on p.**??** and the preceding remarks on structure compatibility. □

The proof of item C in the previous remark is given in the next proposition.

Proposition 11.6 (Linear mappings preserve linear combinations).

Let V_1, V_2 be two vector spaces. Let $f(\cdot) : V_1 \to V_2$ be a linear map and let $x_1, x_2, x_3, \dots, x_n \in V_1$ be a finite number of vectors in the domain V_1 of $f(\cdot)$. Let $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n \in \mathbb{R}$.

Then $f(\cdot)$ preserves any such linear combination, i.e.,

$$f\Big(\sum_{j=0}^{n} \lambda_j x_j\Big) \;=\; \sum_{j=0}^{n} \lambda_j f(x_j). \tag{11.14}$$

PROOF by induction on n: We first note that $f(\lambda x) = \lambda f(x)$ because linear mappings preserve scalar multiples. This proves the base case $n = 1$. Because linear mappings also preserve the addition of any two vectors, the proposition holds for $n = 2$. Our induction assumption is

$$f\Big(\sum_{j=0}^{k} \lambda_j x_j\Big) \;=\; \sum_{j=0}^{k} \lambda_j f(x_j) \quad \text{for all } 1 \le k < n.$$

We use it in the second equation ($k = 2$) and the third equation ($k = n - 1$) of the following:

$$f\Big(\sum_{j=0}^{n} \lambda_j x_j\Big) \;=\; f\Big(\sum_{j=0}^{n-1} \lambda_j x_j + \lambda_n x_n\Big) = \; f\Big(\sum_{j=0}^{n-1} \lambda_j x_j\Big) + f(\lambda_n x_n) = \; \sum_{j=0}^{n-1} \lambda_j f(x_j) + f(\lambda_n x_n) = \; \sum_{j=0}^{n} \lambda_j f(x_j)$$

■

Here are some examples of linear mappings.

Example 11.18 (Projection on the first coordinate). Let $n \in \mathbb{N}$. The map

$$\pi_1(\cdot) : \mathbb{R}^n \to \mathbb{R} \qquad (x_1, x_2, \dots, x_n) \mapsto x_1$$

is called the **projection** on the first coordinate or the first coordinate function. □

Example 11.19 (Projections on any coordinate)**.** More generally, let $n \in \mathbb{N}$ and $1 \leq j \leq n$.

$$\pi_j(\cdot) : \mathbb{R}^n \to \mathbb{R} \qquad (x_1, x_2, \dots, x_n) \mapsto x_j$$

is called the **projection** on the jth coordinate or the **jth coordinate function**.
A specific example for $n = 2$: Let $\vec{v} := (3.5, -2) \in \mathbb{R}^2$. Then $\pi_1(\vec{v}) = 3.5$ and $\pi_2(\vec{v}) = -2$. □

Example 11.20 (Projections on any lower dimensional space)**.** In the last two examples we projected $\mathbb{R}^n$ onto a one–dimensional space. More generally, we can project $\mathbb{R}^n$ onto a vector space $\mathbb{R}^m$ of lower dimension m (i.e., we assume $m < n$) by keeping m of the coordinates and throwing away the remaining $n - m$ coordinates. Mathematicians express this as follows:
Let $m, n, i_1, i_2, \dots, i_m \in \mathbb{N}$ such that $m < n$ and $1 \leq i_1 < i_2 < \cdots < i_m \leq n$. The map

$$\pi_{i_1, i_2, \dots, i_m}(\cdot) : \mathbb{R}^n \to \mathbb{R}^m \qquad (x_1, x_2, \dots, x_n) \mapsto (x_{i_1}, x_{i_2}, \dots, x_{i_m}) \tag{11.15}$$

is called the **projection** on the coordinates $i_1, i_2, \dots, i_m$. [6] □

Example 11.21. Let $x_0 \in A$. The mapping

$$\varepsilon_{x_0} : \mathscr{F}(A, \mathbb{R}) \to \mathbb{R}; \qquad f(\cdot) \mapsto f(x_0) \tag{11.16}$$

which assigns to any real–valued function on A its value at the specific point x_0 is linear because if $h(\cdot) = \sum\limits_{j=0}^{n} a_j f_j(\cdot)$ then

$$\varepsilon_{x_0}\Big(\sum_{j=0}^{n} a_j f_j\Big) = \varepsilon_{x_0}(h) = h(x_0) = \sum_{j=0}^{n} a_j f_j(x_0) = \sum_{j=0}^{n} a_j \varepsilon_{x_0}(f_j).$$

$\varepsilon_{x_0}(\cdot)$ is called the **abstract integral** with respect to point mass at x_0. □

Lemma 11.1 ($F \circ$ *span* $=$ *span* $\circ F$)**.**

Let V, W be two vector spaces and $F : V \to W$ a linear mapping from V to W. Let $A \subseteq V$. Then

$$F(\mathit{span}(A)) = \mathit{span}(F(A)). \tag{11.17}$$

Proof: See Brin/Marchesi Linear Algebra, general lemma 4.1.7. ■

Definition 11.10 (Linear dependence and independence)**.**

[6] You previously encountered an example where we made use of the projection

$$\pi_{1,2}(\cdot) : \mathbb{R}^3 \to \mathbb{R}^2 \qquad (x, y, z) \mapsto (x, y).$$

This was in the course of computing the length of a 3–dimensional vector (see (11.5) on p.506).

★

Let V be a vector space and $A \subseteq V$

(a) A is called **linearly dependent** if the following is true: There exist distinct vectors $x_1, x_2, \dots x_k \in A$ and scalars $\alpha_1, \alpha_2, \dots \alpha_k \in \mathbb{R}$ $(k \in \mathbb{N})$ such that

- not all scalars α_j are zero $(1 \leq j \leq k)$
- $\sum_{j=1}^{k} \alpha_j x_j = 0.$

(b) A is called **linearly independent** if A is not linearly dependent, i.e., if the following is true: Let $x_1, x_2, \dots x_k \in A$ and $\alpha_1, \alpha_2, \dots \alpha_k \in \mathbb{R}$ $(k \in \mathbb{N})$.

- If $\sum_{j=1}^{k} \alpha_j x_j = 0$ then $\alpha_j = 0$, for all $1 \leq j \leq k$. □

Definition 11.11 (Basis of a vector space)**.**

★

Let V be a vector space and $B \subseteq V$. B is called a **basis** of V if both

- B is linearly independent
- *span*$(B) = V$. □

For the Kronecker delta which appears in the next item, see Definition **??** on p.**??**.

Definition 11.12 (Standard basis of $\mathbb{R}^n$)**.**

★

Let $n \in \mathbb{N}$. For $i \in [1, n]_{\mathbb{Z}}$, let $\vec{e}^{(i)} := \big(\delta_{i1}, \delta_{i2}, \dots, \delta_{in}\big)^\top$.

Here δ_{ij} denotes the Kronecker delta: $\delta_{ii} = 1$ for all i and $\delta_{ij} = 0$ for $i \neq j$. Thus,

$$\vec{e}^{(1)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \vec{e}^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \vec{e}^{(n)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Let $B := \{\vec{e}^{(i)} : i \in [1, n]_{\mathbb{Z}}\}$. Then B is a basis of $\mathbb{R}^n$ which we call the **standard basis**, also the **canonical basis**, of $\mathbb{R}^n$. □

Remark 11.7.

If $\vec{x} = \big(x_1, \dots, x_n\big)^\top$, then

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i \vec{e}^{(i)}.$$

From this it is immediate that

- The left hand side of the above is the zero vector if and only if $x_1 = x_2 = \cdots = x_n = 0$. Thus, the standard basis is a set of linearly independent vectors.
- Every $\vec{x} \in \mathbb{R}^n$ is a linear combination of $\vec{e}^{(1)}, \ldots, \vec{e}^{(n)}$. Thus, the linear span of B is $\mathbb{R}^n$.

It follows that the standard basis is, in fact, a basis of $\mathbb{R}^n$. $\square$

Lemma 11.2.

Let V be a vector space and $A \subseteq V$.
Assume that A is linearly independent but not a basis and that $y \in \ span(A)^{\complement}$.
Then $A \cup \{y\}$ is linearly independent.

Proof: Let $A' := A \cup \{y\}$ and let $x_1, x_2, \ldots x_k$ be distinct elements of A' and $\alpha_1, \alpha_2, \ldots \alpha_k \in \mathbb{R}$ $(k \in \mathbb{N})$ such that

$$\sum_{j=1}^{k} \alpha_j x_j \ = \ 0 \tag{11.18}$$

We must show that each α_j is zero.
Case 1: $y \neq x_j$ for all j:
Then $\{x_1, \ldots, x_k\} \subseteq A' \setminus \{y\} = A$. It follows from the linear independence of A that each α_j is zero.
Case 2: $y = x_{j_0}$ for some $1 \leq j_0 \leq k$:
We first show that $\alpha_{j_0} = 0$. This is true because otherwise

$$x_{j_0} \ = \ \sum_{j \neq j_0} \frac{-\alpha_j}{\alpha_{j_0}} x_j \tag{11.19}$$

would be a linear combination of elements of A, contrary to the assumption that $x_{j_0} = y \in \ span(A)^{\complement}$.
We deduce from (11.18) and $\alpha_{j_0} = 0$ that

$$\sum_{j \neq j_0} \alpha_j x_j \ = \ 0. \tag{11.20}$$

It follows from $\{x_j : j \neq j_0\} \subseteq A' \setminus \{y\} = A$ and the linear independence of A that $\alpha_j = 0$ for all $j \neq j_0$. $\blacksquare$

Theorem 11.2.

Let V be a vector space with a finite basis $B = \{b_1, \ldots, b_k\}$.
Then any other basis of V has the same size k.

PROOF:
See, e.g., [1] Brin/Marchesi Linear Algebra. $\blacksquare$

This last theorem gives rise to the following definition.

Definition 11.13 (Dimension of vector spaces).

⋆

- Let V be a vector space with a finite basis $B = \{b_1, \dots, b_k\}$. We call k the **dimension** of V and we write $\dim(V) = k$.
- If V does not possess a finite basis then we say that V has infinite dimension and we write $\dim(V) = \infty$. □

The following proposition gives an example of an infinite linearly independent set.

Proposition 11.7.

For $a \in \mathbb{R}$ define $f_a(\cdot) \in \mathscr{B}(\mathbb{R}, \mathbb{R})$ as follows.

$$f_a(x) \ := \begin{cases} 0 & if x \neq a, \\ 1 & if x = a. \end{cases}$$

Then $\mathscr{A} := \ \{f_a : a \in \mathbb{R}\}$ is a linearly independent subset of $\mathscr{B}(\mathbb{R}, \mathbb{R})$.

PROOF:
We write $0(\cdot)$ for the zero function on $\mathbb{R}$. Let $n \in \mathbb{N}, a_1, \dots a_n \in \mathbb{R}$, and $\alpha_1, \dots \alpha_n \in \mathbb{R}$ such that

$$f \ := \sum_{j=1}^{k} \alpha_j f_{a_j} \ = \ 0(\cdot), \text{ i.e., } \ f(x) = \sum_{j=1}^{k} \alpha_j f_{a_j}(x) \ = \ 0 \text{ for all } x \in \mathbb{R}.$$

We must show that then $\alpha_i = 0$ for all integers $1 \leq i \leq k$. We have $0 = f(a_i) = \sum_{j=1}^{k} f_{a_j}(a_i) \ = \ \alpha_i$.
■

Proposition 11.8.

Let V be a vector space and let U be a (linear) subspace of V. Let $x_0 \in V$.
Let $\tilde{U} \ := \ \{u + \lambda x_0 : u \in U$ and $\lambda \in \mathbb{R}\}$. Then $\tilde{U} = span(U \cup \{x_0\})$.

PROOF:
PROOF of $\subseteq$): Let $x \in \tilde{U}$, i.e., $x = u + \lambda x_0$ for some $u \in U$ and $\lambda \in \mathbb{R}$. Clearly x is a linear combination of $u \in U$ and x_0, hence $x \in span(U \cup \{x_0\})$.
PROOF of $\supseteq$): Let $x \in span(U \cup \{x_0\})$. By the definition of spans there exists $k \in \mathbb{N}$, $u_1, \dots, u_k \in U$ and $\alpha, \alpha_1, \dots \alpha_k \in \mathbb{R}$ such that $x \ = \ \sum_{j=1}^{k} \alpha_j u_j \ + \ \alpha x_0$. Let $u \ := \ \sum_{j=1}^{k} \alpha_j u_j$. Then $x = u + \alpha x_0$, hence $x \in \tilde{U}$. ■

Proposition 11.9.

Let V and V' be two vector spaces and let U be a proper (linear) subspace of V, i.e., $U \subsetneq V$. Let $x_0 \in U^{\complement}$, $y_0 \in V'$. Let $f := U \to V'$ be a linear function from U into V'. Let $\alpha \in \mathbb{R}$. Then

$$g : U \biguplus \{x_0\} \to V'; \quad g(x) := \begin{cases} f(x) & if x \in U, \\ y_0 & if x = x_0, \end{cases} \tag{11.21}$$

uniquely extends to a linear function $\tilde{f}$:span$(U \biguplus \{x_0\}) \to V'$ as follows:

$$\tilde{f}(x + \alpha\, x_0) := f(x) + \alpha\, y_0 \quad for\ x \in U, \alpha \in \mathbb{R}\,. \tag{11.22}$$

PROOF:

Let $\tilde{U} :=$span$(U \biguplus \{x_0\})$. It follows from prop.11.8 on p.521 that any $x \in \tilde{U}$ is of the form $x = u + \alpha x_0$ for some suitable $u \in U$ and $\alpha \in \mathbb{R}$.

It follows that the function $\tilde{f}$ defined in (11.22) is in fact defined on all of $\tilde{U}$. Clearly $\tilde{f}$ coincides with g on $U \biguplus \{x_0\}$, hence $\tilde{f}$ extends g from $U \biguplus \{x_0\}$ to $\tilde{U}$.

Proof of linearity of $\tilde{f}$:

Let x_1 and $x_2 \in \tilde{U}$, i.e., there exist $u_1, u_2 \in U$ and $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $x_1 = u_1 + \alpha_1 x_0$ and $x_2 = u_2 + \alpha_2 x_0$. Let $\lambda \in \mathbb{R}$. To prove linearity of $\tilde{f}$ we must show that $\tilde{f}(x_1 + \lambda x_2) = \tilde{f}(x_1) + \lambda \tilde{f}(x_2)$.

$$\begin{aligned} &\tilde{f}(x_1 + \lambda x_2) = \tilde{f}\big((u_1 + \alpha_1 x_0) + \lambda(u_2 + \alpha_2 x_0)\big) = \tilde{f}\big((u_1 + \lambda u_2) + (\alpha_1 x_0 + \lambda \alpha_2 x_0)\big) \\ &= \tilde{f}\big((u_1 + \lambda u_2) + (\alpha_1 + \lambda \alpha_2) x_0\big) = f(u_1 + \lambda u_2) + (\alpha_1 + \lambda \alpha_2) y_0 \\ &= \big(f(u_1) + \lambda f(u_2)\big) + (\alpha_1 y_0 + \lambda \alpha_2 y_0) = \big(f(u_1) + \alpha_1 y_0\big) + \lambda\big(f(u_2) + \alpha_2 y_0\big) = \tilde{f}(x_1) + \lambda \tilde{f}(x_2). \end{aligned}$$

The linearity of f on U was used in the fifth equation. Everything else is utilizing (11.22) and grouping terms differently. This finishes the proof of linearity of $\tilde{f}$ on $\tilde{U}$.

It remains to show the uniqueness of $\tilde{f}$. So let $h : \tilde{U} \to V'$ linear such that $h(x) = \tilde{f}(x)$ for all $x \in U \biguplus \{x_0\}$. We must prove that $h(x) = \tilde{f}(x)$ for all $x \in \tilde{U}$. Let $x \in \tilde{U}$, i.e., $x = u + \alpha x_0$ for some $u \in U$ and $\alpha \in \mathbb{R}$. Then

$$h(x) = h(u + \alpha x_0) = h(u) + \alpha h(x_0) = \tilde{f}(u) + \alpha \tilde{f}(x_0) = \tilde{f}(u + \alpha x_0) = \tilde{f}(x).$$

The third equality results from $h\Big|_{U \biguplus \{x_0\}} = \tilde{f}\Big|_{U \biguplus \{x_0\}}$, the second and fourth equalities from the linearity of h and $\tilde{f}$. This proves uniqueness of $\tilde{f}$. ■

11.2.2 Normed Vector Spaces

Definition 11.4 on p.507 in ch.11.1.3 (Length of n–Dimensional Vectors and the Euclidean Norm) gave the definition of the Euclidean norm $\|\vec{x}\|_2 = \sqrt{\sum_{j=1}^{n} x_j^2}$ in $\mathbb{R}_n$. We saw that in dimensions $n = 1, 2, 3$ that $\|\vec{x}\|_2$ equals the length of the vector $\vec{x}$ and that prop.11.1 on p. 508 "proved" informally for $n = 1, 2, 3$ that $\|\cdot\|_2$ satisfies the following three properties:

(a) positive definiteness,
(b) absolute homogeneity,
(c) triangle inequality.

In this chapter we define the norm $\|x\|$ of a vector x in an abstract vector space as a function which satisfies the above three properties, and hence generalizes the concept of the length of a vector in n–dimensional space to more general vector spaces. Before we give give this definition, we first introduce the concept of an inner product $x \bullet y$ of two vectors x and y. We will see that some of the most important norms, the Euclidean norm among them, can be derived from inner products.

The following definition of inner products and proof of the Cauchy–Schwartz inequality were taken from "Calculus of Vector Functions" (Williamson/Crowell/Trotter [4]).

Definition 11.14 (Inner product)**.**

Let V be a vector space with a function

$$\bullet(\cdot,\cdot) : V \times V \to \mathbb{R}; \qquad (x,y) \mapsto x \bullet y := \bullet(x,y)$$

which satisfies the following:

(11.23a) $x \bullet x \geq 0 \quad \forall x \in V \quad \text{and} \quad x \bullet x = 0 \ \Leftrightarrow \ x = 0$ **positive definiteness**

(11.23b) $x \bullet y = y \bullet x \quad \forall x, y \in V$ **symmetry**

(11.23c) $(x+y) \bullet z = x \bullet z + y \bullet z \quad \forall\, x, y, z \in V$ **additivity**

(11.23d) $(\lambda x) \bullet y = \lambda (x \bullet y) \quad \forall\, x, y \in V \quad \forall\, \lambda \in \mathbb{R}$ **homogeneity**

We call such a function an **inner product**. ☐

An inner product is also referred to as a **dot product**, e.g., in [1] Brin/Marchesi Linear Algebra, ch.6, Orthogonality.

Also note that additivity and homogeneity of the mapping $x \mapsto x \bullet y$ for a fixed $y \in V$ imply linearity of that mapping and the symmetry property implies that the mapping $y \mapsto x \bullet y$ for a fixed $x \in V$ is linear too. In other words, an inner product is binear in the following sense:

Definition 11.15 (Bilinearity)**.**

⋆

Let V be a vector space with a function

$$B : V \times V \to \mathbb{R}; \qquad (x,y) \mapsto B(x,y).$$

$B(\cdot,\cdot)$ is called **bilinear** if it is linear in each argument, i.e., the mappings

$$B_1 : V \to \mathbb{R}; \quad x \mapsto B(x,y)$$
$$B_2 : V \to \mathbb{R}; \quad y \mapsto B(x,y)$$

are both linear. ☐

Proposition 11.10 (Algebraic properties of the inner product)**.**

Let V be a vector space with inner product $\bullet(\cdot,\cdot)$. Let $a, b, x, y \in V$. Then

(11.24a) $$(a+b) \bullet (x+y) = a \bullet x \ + \ b \bullet x \ + \ a \bullet y \ + \ b \bullet y$$

(11.24b) $$(x+y) \bullet (x+y) = x \bullet x \ + \ 2(x \bullet y) \ + \ y \bullet y$$

(11.24c) $$(x-y) \bullet (x-y) = x \bullet x \ - \ 2(x \bullet y) \ + \ y \bullet y$$

PROOF of **(b)** and **(c)**: Left as an exercise.
PROOF of **a:**

$$\begin{aligned}(a+b) \bullet (x+y) &= (a+b) \bullet x \ + \ (a+b) \bullet y \\ &= a \bullet x \ + \ b \bullet x \ + a \bullet y \ + \ b \bullet y.\end{aligned}$$

We used linearity in the second argument for the first equality and linearity in the first argument for the second equality.
The proof of **(b)** and **(c)** is left as exercise 11.2 (see p.540). ■

The following is the most important example of an inner product.

Proposition 11.11 (Inner product on $\mathbb{R}^n$))**.**

Let $n \in \mathbb{N}$. Then the real–valued function

(11.25) $$(\vec{x}, \vec{y}) \ \mapsto \ x_1y_1 + x_2y_2 + \ldots + x_ny_n \ = \ \sum_{j=1}^{n} x_jy_j \,,$$

where $\vec{x} = (x_1, \ldots x_n)$ and $\vec{y} = (y_1, \ldots y_n)$, is an inner product on $\mathbb{R}^n \times \mathbb{R}^n$.

PROOF:

(a) For $\vec{x} = \vec{y}$ we obtain $\vec{x} \bullet \vec{x} \ = \ \sum_{j=1}^{n} x_j^2$ and positive definiteness of the inner product follows from $\sum_{j=1}^{n} x_j^2 = 0 \ \Leftrightarrow \ x_j^2 = 0 \ \forall j \ \Leftrightarrow \ x_j = 0 \ \forall j$.

(b) Symmetry is clear because $x_jy_j = y_jx_j$.

(c) Let $\vec{z} = (z_1, \ldots z_n)$. Additivity follows from the fact that $(x_j + y_j)z_j = x_jz_j + y_jz_j$.

(d) Homogeneity follows from the fact that $(\lambda x_j)y_j = \lambda(x_jy_j)$. ■

Proposition 11.12 (Cauchy–Schwartz inequality for inner products)**.**

> *Let V be a vector space with an inner product*
> $$\bullet(\cdot,\cdot) : V \times V \to \mathbb{R}; \qquad (x,y) \mapsto x \bullet y := \bullet(x,y)$$
> *Then,*
> $$(x \bullet y)^2 \;\le\; (x \bullet x)\;(y \bullet y).$$

PROOF:
Step 1: We assume first that $x \bullet x = y \bullet y = 1$. Then

$$\begin{aligned} 0 \;&\le (x-y) \bullet (x-y) \\ &= x \bullet x - 2x \bullet y + y \bullet y = 2 - 2(x \bullet y) \end{aligned}$$

where the first equality follows from proposition (11.10) on p.524. Thus $2(x \bullet y) \;\le\; 2$, i.e.,

(11.26) $$x \bullet y \;\le\; 1.$$

Since this inequality holds for any vectors x, y such that $x \bullet x = y \bullet y = 1$ and since absolute homogeneity (11.23d) implies $(-x) \bullet (-x) = (-1)^2 x \bullet x = 1$ we may replace x with $-x$ and obtain

(11.27) $$-(x \bullet y) \;=\; (-x) \bullet y \;\le\; 1.$$

It follows from (11.26) and (11.27) that $|x \bullet y| \;\le\; 1$, thus $(x \bullet y)^2 \;\le\; 1$, i.e., $(x \bullet y)^2 \;\le\; (x \bullet x)\;(y \bullet y)$. The Cauchy–Schwartz inequality is thus true under the assumption $x \bullet x = y \bullet y = 1$.

Step 2: General case: We do not assume anymore that $x \bullet x = y \bullet y = 1$. If x or y is zero then the Cauchy–Schwartz inequality is trivially true because, say if $x = 0$ then the left–hand side becomes

$$(x \bullet y)^2 = (0x \bullet y)^2 = 0(x \bullet y)^2 = 0$$

whereas the right–hand side is, as the product of two nonnegative numbers $x \bullet x$ and $y \bullet y$, nonnegative.

So we can assume that x and y are not zero. On account of the positive definiteness we have $x \bullet x > 0$ and $y \bullet y > 0$. This allows us to define $u := x/\sqrt{x \bullet x}$ and $v := y/\sqrt{y \bullet y}$. But then

$$\begin{aligned} u \bullet u \;&= (x \bullet x)/\sqrt{x \bullet x}^2 \;= 1 \\ v \bullet v \;&= (y \bullet y)/\sqrt{y \bullet y}^2 \;= 1. \end{aligned}$$

We have already seen in step 1 that $u \bullet v \;\le\; 1$. It follows that

$$(x \bullet y)/(\sqrt{x \bullet x}\sqrt{y \bullet y}) \;=\; (x/\sqrt{x \bullet x}) \bullet (y/\sqrt{y \bullet y}) \;\le\; 1$$

We multiply both sides with $\sqrt{x \bullet x}\sqrt{y \bullet y}$,

$$x \bullet y \;\le\; \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

We replace x by $-x$ and obtain

$$-(x \bullet y) \;\le\; \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

Because $|x \bullet y|$ is either of $-(x \bullet y)$ or $(x \bullet y)$, it follows from the last two inequalities that

$$|x \bullet y| \;\le\; \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

We square this and obtain

$$(x \bullet y)^2 \;\leq\; (x \bullet x)\,(y \bullet y)$$

and the Cauchy–Schwartz inequality is proven. ■

Definition 11.16 (sup–norm of bounded real–valued functions)**.**

Let X be an arbitrary, nonempty set. Let $f : X \to \mathbb{R}$ be a bounded real–valued function on X, i.e., there exists $K \geq 0$ such that $|f(x)| \leq K$ for all $x \in X$. Let

$$\|f\|_\infty \;:=\; \sup\{|f(x)| : x \in X\} \tag{11.28}$$

We call $\|f\|_\infty$ the **supremum norm** or **sup–norm** of the function f. □

Proposition 11.13 (Properties of the sup norm)**.**

Let X be an arbitrary, nonempty set. Let

$$\mathscr{B}(X, \mathbb{R}) := \{h(\cdot) : h(\cdot) \text{ is a bounded real–valued function on } X\}$$

(see example 11.12 on p. 513). Then the sup–norm

$$\|\cdot\|_\infty : \mathscr{B}(X,\mathbb{R}) \to \mathbb{R}_+, \qquad h \mapsto \|h\|_\infty = \sup\{|h(x)| : x \in X\}$$

satisfies the following:

$$\|f\|_\infty \geq 0 \;\forall\; f \in \mathscr{B}(X,\mathbb{R}) \text{ and } \|f\|_\infty = 0 \;\Leftrightarrow\; f(\cdot) = 0 \qquad \textbf{positive definiteness} \tag{11.29a}$$

$$\|\alpha f(\cdot)\|_\infty = |\alpha| \cdot \|f(\cdot)\|_\infty \;\forall\; f \in \mathscr{B}(X,\mathbb{R}), \forall\; \alpha \in \mathbb{R} \qquad \textbf{absolute homogeneity} \tag{11.29b}$$

$$\|f(\cdot) + g(\cdot)\|_\infty \leq \|f(\cdot)\|_\infty + \|g(\cdot)\|_\infty \;\forall\; f, g \in \mathscr{B}(X,\mathbb{R}) \qquad \textbf{triangle inequality} \tag{11.29c}$$

PROOF: The proof is left as exercise 11.1 on p.540. ■

Note 11.2. We previously discussed the Euclidean norm

$$\|\vec{x}\|_2 \;=\; \sqrt{\sum_{j=1}^{n} {x_j}^2} \tag{11.30}$$

for n–dimensional vectors $\vec{x} = (x_1, x_2, \dots, x_n)$. You saw in (11.1) on p.508 that it satisfies positive definiteness, absolute homogeneity and the triangle inequality, just like the sup–norm. [7] Those are properties which you associate with the length or size of an object. A very rich mathematical theory can be developed for a generalized definition of length which is based just on those properties. □

[7] Actually, the proof that $\|\cdot\|_2$ satisfies the triangle inequality was given only for dimensions $1, 2, 3$. It will be proved in this chapter that it is true for all dimensions n. See cor.11.1 (Inner products define norms) on p.529.

As mentioned before, mathematicians like to define new objects that are characterized by a certain set of properties. As an example we had the definition of a vector space which encompasses objects as different as finite–dimensional vectors and real–valued functions. Accordingly we give a special name to a function defined on a vector space which satisfies positive definiteness, homogeneity and the triangle inequality.

Definition 11.17 (Normed vector spaces).

Let V be a vector space with a real–valued function

$$\|\cdot\| : \quad V \to \mathbb{R} \qquad x \mapsto \|x\|$$

which satisfies

(11.31a)	$\|x\| \geq 0 \quad \forall x \in V \quad \text{and} \quad \|x\| = 0 \;\Leftrightarrow\; x = 0$	**positive definiteness**
(11.31b)	$\|\alpha x\| = \|\alpha\| \cdot \|x\| \quad \forall x \in V, \forall \alpha \in \mathbb{R}$	**absolute homogeneity**
(11.31c)	$\|x + y\| \leq \|x\| + \|y\| \quad \forall\, x, y \in V$	**triangle inequality**

We call $\|\cdot\|$ a **norm** on V and we call V a **normed vector space**.

We write $(V, \|\cdot\|)$ instead of V when we wish to emphasize what norm on V we are discussing. □

Example 11.22 (Vector space of polynomials with sup–norm).

Let $A \subseteq \mathbb{R}$. It follows from (**??**) and (**??**) that the set $\mathscr{P} := \{p(\cdot) : p(\cdot) \text{ is a polynomial on } A\}$ of all polynomials on an arbitrary nonempty subset A of the real numbers is a subspace of the vector space $\mathscr{C}(A, \mathbb{R})$. (see example (**??**) on p.**??**.

If A is bounded then any polynomial p on A is bounded, hence its sup–norm

$$\|p\|_\infty \;=\; \sup\{|p(x)| : x \in A\}$$

is finite, and $(\mathscr{P}, \|\cdot\|_\infty)$ is a normed vector space.

If A is not bounded, then $\|p\|_\infty$ is not finite for all $p \in \mathscr{P}$. Matter of fact, it can be shown that, if A is not bounded, then the only polynomials for which $\|p(\cdot)\|_\infty < \infty$ are the constant functions on A. □

Remark 11.8. Let $(V, \|\cdot\|)$ be a normed vector space and let $\gamma > 0$.
Let $p : V \to \mathbb{R}$ be defined as $p(x) := \gamma\|x\|$. Then p also is a norm.

PROOF: The proof is left as exercise 11.3. ■

Definition 11.18 (p–norms for $\mathbb{R}^n$).

⋆

Let $p \geq 1$. It will be proved in prop.11.16 on p.533 that the function

$$\vec{x} \mapsto \|\vec{x}\|_p := \Big(\sum_{j=1}^{n} |x_j|^p\Big)^{1/p} \tag{11.32}$$

is a norm on $\mathbb{R}^n$). This norm is called the **p–norm** on $\mathbb{R}^n$). The Euclidean norm is a p–norm; it is the 2-norm on $\mathbb{R}^n$). ☐

Remark 11.9.

We have seen that a vector space can be endowed with more than one norm.

(a) We have seen in Remark11.8 that if $x \mapsto \|x\|$ is a norm on a vector space V and $\beta > 0$ then $x \mapsto \beta \cdot \|x\|$ also is a norm on V.

(b) The p–norms define a collection of different norms for $\mathbb{R}^n$. ☐

The following theorem shows that an inner product can be associated in a natural fashion with a norm.

Theorem 11.3 (Inner products define norms)**.**

Let V be a vector space with an inner product

$$\bullet(\cdot,\cdot) : V \times V \to \mathbb{R}; \qquad (x,y) \mapsto x \bullet y$$

Then

$$\|\cdot\|_\bullet : x \mapsto \|x\| = \sqrt{(x \bullet x)}$$

defines a norm on V

PROOF:

Positive definiteness : follows immediately from that of the inner product.

Absolute homogeneity : Let $x \in V$ and $\lambda \in \mathbb{R}$. Then

$$\|\lambda x\|_\bullet = \sqrt{(\lambda x) \bullet (\lambda x)} = \sqrt{\lambda\lambda(x \bullet x))} = |\lambda|\sqrt{x \bullet x} = |\lambda|\|x\|_\bullet.$$

Triangle inequality : Let $x, y \in V$. Then

$$\begin{aligned}\|x+y\|_\bullet{}^2 &= (x+y) \bullet (x+y)\\ &= x \bullet x + 2(x \bullet y) + y \bullet y\\ &\leq x \bullet x + 2|x \bullet y| + y \bullet y\\ &\leq x \bullet x + 2\sqrt{x \bullet x}\sqrt{y \bullet y} + y \bullet y\\ &= \|x\|_\bullet{}^2 + 2\|x\|_\bullet \|y\|_\bullet + \|y\|_\bullet{}^2\\ &= \big(\|x\|_\bullet + \|y\|_\bullet\big)^2.\end{aligned}$$

The second equation uses bilinearity and symmetry of the inner product. The first inequality expresses the simple fact that $\alpha \leq |\alpha|$ for any number α. The second inequality uses Cauchy–Schwartz. The next equality just substitutes the definition $\|x\|_\bullet = \sqrt{(x \bullet x)}$ of the norm. The next and last equality is the binomial expansion $(a+b)^2 = a^2 + 2ab + b^2$ for the ordinary real numbers $a = \|x\|_\bullet$ and $b = \|y\|_\bullet$.

We take square roots in the above inequality $\|x+y\|_\bullet{}^2 \leq (\|x\|_\bullet + \|y\|_\bullet)^2$ and obtain $\|x+y\|_\bullet \leq \|x\|_\bullet + \|y\|_\bullet$, the triangle inequality we set out to prove. ■

Definition 11.19 (Norm for an inner product)**.**

Let V be a vector space with an inner product

$$\bullet(\cdot,\cdot) : V \times V \to \mathbb{R}; \qquad (x,y) \mapsto x \bullet y$$

Then

$$\|\cdot\|_\bullet : x \ \mapsto\ \|x\|_\bullet \ :=\ \sqrt{(x \bullet x)} \tag{11.33}$$

is called the **norm associated with the inner product** $\bullet(\cdot,\cdot)$. □

It was stated in prop.11.1 on p. 508 that the Euclidean norm is in fact a norm but only positive definiteness and homogeneity were proved. We now can easily complete the proof.

Corollary 11.1.

The Euclidean norm in $\mathbb{R}^n$*:*

$\|(x_1, x_2, \dots, x_n)\|_2 \ = \sqrt{\sum_{j=1}^{n} {x_j}^2}$ *(see def.11.4 on p.507) is a norm.*

PROOF: This follows from the fact that

$$\vec{x} \bullet \vec{y} \ = \sum_{j=1}^{n} x_j y_j \qquad \text{where } \vec{x} = (x_1, \dots, x_n) \text{ and } \vec{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$$

defines an inner product on $\mathbb{R}^n \times \mathbb{R}^n$ (see prop.11.11) for which $\|(x_1, x_2, \dots, x_n)\|_2$ is the associated norm. ■

We now look at an inner product on the vector space $\mathscr{C}([a,b],\mathbb{R})$ of all continuous real–valued functions on the interval $[a,b]$ which was defined in example 11.12 (Vector spaces of real–valued functions) on p.513. We use the terminology of [3] Stewart, J: Single Variable Calculus) for the following.

Definition 11.20.

★ Let $a, b \in \mathbb{R}$, $a < b$ and assume that $f, g : [a, b] \to \mathbb{R}$ are integrable functions. (See example **??** on p.**??**.)

(a) We call the definite integral $\int_a^b f(x)dx$ the **net area** between the graph of f, the x–axis, and the vertical lines through $(a, 0)$ $(y = a)$ and $(b, 0)$ $(y = b)$. The above integral treats areas above the x–axis as positive and below the x–axis as negative, i.e., the net area is the difference between the areas above the x–axis and those below the x–axis.

(b) We call $\int_a^b |f(x)|dx$ the **area** between the graph of f, the x–axis, and the vertical lines $y = a$ and $y = b$. Note that $f(x)$ has been replaced by its absolute value $|f(x)|$. In contrast to the net area, areas below the x–axis are also counted positive. ☐

(c) We call $\int_a^b f(x) - g(x)dx$ the **net area** between the graphs of f and g and the vertical lines $y = a$ and $y = b$. We call $\int_a^b |f(x) - g(x)|dx$ the **area** between the graphs of f and g and the vertical lines $y = a$ and $y = b$. ☐

Example 11.23. Let $f : [-1, 1];\ x \mapsto 4x^3$. The antiderivative (see example **??** on p.**??**) of $f(\cdot)$ is $x \mapsto x^4$ and we compute net area and area as follows:

(a) Net area $= \int_{(-1)}^1 4x^3 dx \ = \ x^4\Big|_{-1}^1 \ = \ 1 - 1 = 0;$

(b) Area
$$= \int_{(-1)}^1 4|x^3|dx \ = \ \int_{(-1)}^0 (-4x^3)dx \ + \ \int_0^1 4x^3 dx$$
$$= \ -x^4\Big|_{-1}^0 \ + \ -x^4\Big|_0^1 \ = \ \big(0 - (-1)\big) \ + \ \big(1 - 0\big) = 2. \quad \square$$

Let $a, b \in \mathbb{R}$ such that $a < b$. We remember from example **??** on p.**??** that continuous functions are integrable. This allows us to compare for $f \in \mathscr{C}([a, b], \mathbb{R})$ the expressions

$$\|f\|_\infty \ = \ \sup\{|f(x)| : x \in X\}, \quad \int_a^b |f(x)|dx, \ \text{ and } \ \int_a^b \big(f(x)\big)^2 dx. \tag{11.34}$$

All three expressions give in a sense the size of f. The sup–norm measures it as the biggest possible displacement from zero, the integral over the absolute value measures the area between the graphs of the functions $x \mapsto f(x)$ and $x \mapsto 0$, and the last expression does the same with the square of f. In many respects the use of areas is superior to using the biggest difference to zero.

Squaring $f(\cdot)$ rather than using its absolute value has some mathematical advantages. One of them is that the function $(f, g) \mapsto \int_a^b f(x)g(x)dx$ defines an inner product on $\mathscr{C}([a, b], \mathbb{R})$ whose associated norm is $f \mapsto \int_a^b \big(f(x)\big)^2 dx$. We will discuss that now. In preparation we prove the following proposition.

Proposition 11.14.

Let $a, b \in \mathbb{R}$ such that $a < b$. and let $f : [a, b] \to \mathbb{R}$ be continuous. Then

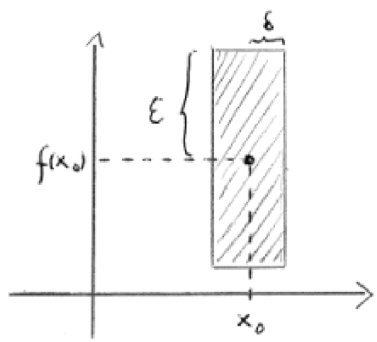
$$\int_a^b f(x)dx \;=\; 0 \quad \textit{if and only if} \quad f(x) \;=\; 0 \;\; \textit{for all } x \in \,]a, b[. \;\; \square$$

PROOF: Assume that there is $a < x_0 < b$ such that $f(x_0) \neq 0$, i.e., $f(x_0) > 0$. Let $\varepsilon := \frac{f(x_0)}{2}$. As f is continuous at x_0 there exists according to thm.?? on p.?? some $\delta > 0$ such that

$$(11.35) \quad |f(x_0) - f(x)| < \varepsilon, \text{ hence } f(x) \;>\; f(x_0) - \varepsilon \;=\; \frac{f(x_0)}{2} \;=\; \varepsilon \text{ for all } x_0 - \delta < x < x_0 + \delta.$$

Continuity at x_0:

If $|x - x_0| < \delta$ then $|f(x_0) - f(x)| < \varepsilon$: The graph of f stays within the rectangle with corners $(x_0 \pm \delta, f(x_0) \pm \varepsilon)$.



Let $g : [a, b] \to \mathbb{R}$ be defined as follows.

$$g(x, y) \;=\; \begin{cases} \varepsilon & \text{if } x_0 - \delta < x < x_0 + \delta \\ 0 & \text{else.} \end{cases}$$

It follows from (11.35) that $f \geq g$, hence $\int_a^b f(x)dx \geq \int_a^b g(x)dx = (2\delta)\varepsilon > 0$. ■

Proposition 11.15.

Let $a, b \in \mathbb{R}$ such that $a < b$. Then the mapping

$$(11.36) \qquad (f, g) \;\mapsto\; f \bullet g \;:=\; \int_a^b f(x)g(x)dx$$

defines an inner product on $f \in \mathscr{C}([a, b], \mathbb{R})$. $\square$

PROOF: We must prove positive definiteness, symmetry, and linearity in the left argument. In the following let $f, g, h \in \mathscr{C}([a, b], \mathbb{R})$ and $\lambda \in \mathbb{R}$.

(a) Positive definiteness: It follows from $f^2(x) \geq 0$ that $f \bullet f \;=\; \int_a^b f^2(x)dx \geq 0$. Clearly, if 0 denotes as usual the zero function $x \mapsto 0$ then $0 \bullet 0 = 0$. It remains to be shown that if $\int_a^b f^2(x)dx \geq 0$ then $f = 0$. This follows from prop.11.14.

(b) Symmetry:

$$f \bullet g \;=\; \int_a^b f(x)g(x)dx \;=\; \int_a^b g(x)f(x)dx \;=\; g \bullet f.$$

(c) Additivity and homogeneity: This can be deduced from the well–known formulas

$$\int_a^b f(x) + g(x)dx \;=\; \int_a^b f(x)dx \;+\; \int_a^b g(x)dx \quad \text{and} \quad \int_a^b \lambda g(x)dx \;=\; \lambda \int_a^b g(x)dx.$$

as follows:

$$(f+g) \bullet h \;=\; \int_a^b \big(f(x)+g(x)\big)h(x)dx \;=\; \int_a^b f(x)h(x)dx \;+\; \int_a^b g(x)h(x)dx \;=\; f \bullet h + g \bullet h,$$

$$(\lambda f) \bullet g \;=\; \int_a^b \lambda f(x)g(x)dx \;=\; \lambda \int_a^b f(x)g(x)dx \;=\; \lambda(f \bullet g). \;\blacksquare$$

According to Definition 11.19 (norm for an inner product) and thm.11.3 (inner products define norms) we now define the norm associated with $f \bullet g = \int_a^b f(x)g(x)dx$.

Definition 11.21 (L_2–Norm for continuous functions)**.**

Let $a, b \in \mathbb{R}$ such that $a < b$. Let $f \bullet g$ be the the following inner product on the space $\mathscr{C}([a,b],\mathbb{R})$ of all continuous functions $[a,b] \to \mathbb{R}$:

(11.37) $$f \bullet g \;:=\; \int_a^b f(x)g(x)dx.$$

The L^2**–norm**. of f is the norm associated with that inner product:

(11.38) $$\|\cdot\|_{L^2} :\; f \mapsto \|f\|_\bullet \;=\; \sqrt{\int_a^b f^2(x)dx}\,.$$

□

We saw in Definition 11.18 that the Euclidean norm is the p–norm $\|\vec{x}\|_p \;=\; \Big(\sum_{j=1}^n |x_j|^p\Big)^{1/p}$ for the special case $p = 2$. There is an analogue for the L^2 norm.

Definition 11.22 (L^p–norms for $\mathscr{C}([a,b],\mathbb{R})$)**.**

⋆ Let $a, b \in \mathbb{R}$ such that $a < b$ and $p \geq 1$.
It will be shown in prop.11.17 (The L^p–norm is a norm) on p.533 that

(11.39) $$f \mapsto \|f\|_{L^p} \;:=\; \Big(\int_a^b |f(x)|^p dx\Big)^{1/p}$$

is a norm on $\mathscr{C}([a,b],\mathbb{R})$. This norm is called the L^p**–norm** of f. □

11.2.3 The Inequalities of Young, Hoelder, and Minkowski ⋆

Note that this chapter is starred, hence optional.

Proposition 11.16 (The p–norm in $\mathbb{R}^n$ is a norm)**.**

> *Let* $p \in [1, \infty[$.
>
> *Then the p-norm* $\vec{x} \mapsto \|\vec{x}\|_p = \left(\sum_{j=1}^n |x_j|^p\right)^{1/p}$ *is a norm in* $\mathbb{R}^n$.

PROOF:

(a). Positive definiteness:

Clearly, $\sum_{j=1}^n |x_j|^p \geq 0$ because each term $|x_j|^p$ is nonnegative, hence $\|\vec{x}\|_p = \sqrt{\sum_{j=1}^n |x_j|^p} \geq 0$. Note that $\|\vec{x}\|_p = 0$ is only possible if $|x_j|^p = 0$ for all indices j, because, if $x_{j_0} \neq 0$ for some j_0 then $|x_{j_0}|^p > 0$, hence $\left(\|\vec{x}\|_p\right)^{1/p} \geq |x_{j_0}|^p > 0$.

(b). Absolute homogeneity:

If $\lambda \in \mathbb{R}$ then

$$\|(\lambda\vec{x})\|_p = \Big(\sum_{j=1}^n (|\lambda|\ |x_j|)^p\Big)^{1/p} = \Big(|\lambda|^p \sum_{j=1}^n |x_j|^p\Big)^{1/p} = |\lambda|\Big(\sum_{j=1}^n |x_j|^p\Big)^{1/p} = |\lambda|\|\vec{x}\|_p.$$

(c). Triangle inequality for $p = 1$:

It follows from $|x_j + y_j| \leq |x_j| + |y_j|$ for all j that

$$\|\vec{x} + \vec{y}\|_1 = \sum_{j=1}^n |x_j + y_j| \leq \sum_{j=1}^n |x_j| + \sum_{j=1}^n |y_j| = \|\vec{x}\|_1 + \|\vec{y}\|_1.$$

(d). Triangle inequality for $p > 1$:

This is Minkowski's inequality for $(\mathbb{R}^n, \|\cdot\|_p)$ (thm.11.7 below). That $\|\cdot\|_2$ satisfies the triangle inequality (i.e., $p = 2$) also follows independently from cor.11.1 on p.529. ■

Proposition 11.17 (The L^p–norm is a norm)**.**

> *Let* $p \in [1, \infty[$ *and let* $a, b \in \mathbb{R}$ *such that* $a < b$.
>
> *Then the* L^p*-norm* $f \mapsto \|f\|_{L^p} = \left(\int_a^b |f(x)|^p\right)^{1/p}$ *is a norm in* $\mathscr{C}([a, b], \mathbb{R})$.

PROOF:

(a). Positive definiteness:

Follows from prop.11.14 on p.530 and the fact that $x \mapsto |f(x)|^p$ is a nonnegative and continuous function.

(b). Absolute homogeneity:

If $\lambda \in \mathbb{R}$ then

$$\begin{aligned}\|(\lambda f)\|_{L^p} &= \Big(\int_a^b (|\lambda|\ |f(x)|)^p\ dx\Big)^{1/p} \\ &= \Big(|\lambda|^p \int_a^b |f(x)|^p\ dx\Big)^{1/p} = |\lambda|\Big(\int_a^b |f(x)|^p\ dx\Big)^{1/p} = |\lambda|\|f\|_{L^p}.\end{aligned}$$

(c). Triangle inequality for $p = 1$:
It follows from $|f(x) + g(x)| \le |f(x)| + |g(x)|$ for all x that

$$\begin{aligned}\|f+g\|_{L^1} &= \int_a^b |f(x)+g(x)|\,dx \;\le\; \int_a^b (|f(x)| \;+\; |g(x)|)\,dx \\ &= \int_a^b |f(x)|\,dx \;+\; \int_a^b |g(x)|\,dx \;=\; \|f\|_{L^1} \;+\; \|g\|_{L^1}.\end{aligned}$$

(d). Triangle inequality for $p > 1$:
This is Minkowski's inequality for L^p–norms (thm.11.5 below). That $\|\cdot\|_{L^2}$ satisfies the triangle inequality (i.e., $p = 2$) also follows independently from cor.11.1 on p.529. ■

We were referring to Minkowksi's inequalities for $(\mathbb{R}^n, \|\cdot\|_p)$ and L^p–norms when proving the triangle inequality for those norms. We now build the machinery that will allows us to prove those inequalities.

Proposition 11.18 (Young's Inequality)**.**

Let $a, b > 0$ and let $p, q > 1$ be ***conjugate indices****, i.e.,*

$$\frac{1}{p} \;+\; \frac{1}{q} \;=\; 1. \tag{11.40}$$

Then ***Young's inequality*** *holds:*

$$ab \;\le\; \frac{a^p}{p} \;+\; \frac{b^q}{q}. \tag{11.41}$$

PROOF:
Step 1: We show that $q - 1 = \frac{1}{p-1}$:

$$\begin{aligned}\frac{1}{p} \;+\; \frac{1}{q} \;=\; 1 \;&\Rightarrow q + p \;=\; pq \;\Rightarrow\; q(1-p) \;=\; -p \\ &\Rightarrow q \;=\; \frac{p}{p-1} \Rightarrow\; q - 1 \;=\; \frac{p-(p-1)}{p-1} \;=\; \frac{1}{p-1}.\end{aligned} \tag{11.42}$$

Step 2: The functions

$$\varphi : [0, \infty[\;\to\; [0, \infty[;\;\; x \mapsto x^{p-1} \quad \text{and} \quad \psi : [0, \infty[\;\to\; [0, \infty[;\;\; y \mapsto y^{q-1}$$

are inverse to each other because we have

$$\psi\big(\varphi(x)\big) \;=\; \psi\big(x^{p-1}\big) \;=\; \big(x^{p-1}\big)^{q-1} \;\overset{(\star)}{=}\; \big(x^{p-1}\big)^{1/(p-1)} \;=\; x$$

($(\star)$ follows from step 1). We further have

$$\varphi\big(\psi(y)\big) \;=\; \varphi\big(y^{q-1}\big) \;=\; \big(y^{q-1}\big)^{p-1} \;\overset{(\star\star)}{=}\; \big(y^{q-1}\big)^{1/(q-1)} \;=\; y$$

(($\star\star$) again follows from step 1). Note that those two functions are continuous (actually, differentiable) and strictly increasing because $\varphi'(t) = (p-1)t^{(p-2)} > 0$ and $\psi'(t) = (q-1)t^{(q-2)} > 0$ for all $t \geq 0$. We further have $\varphi(0) = 0 = \psi(0)$.

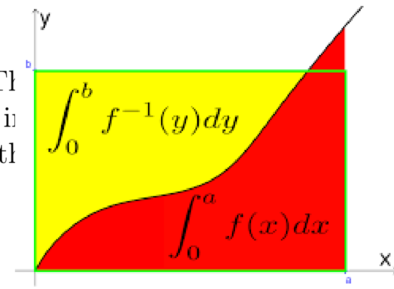
Step 3: Let $f : [0, \infty[\to [0, \infty[$ be a continuous and strictly increasing (hence invertible) function such that $f(0) = 0$. Then the following is true for any two real numbers $a, b > 0$:

$$ab \leq \int_0^a f(x)dx + \int_0^b f^{-1}(y)dy. \tag{11.43}$$

To prove this, we distinguish three cases. Either $b < f(a)$ or $b > f(a)$ or $b = f(a)$.

The picture to the right shows what happens if $b < f(a)$: Tl rectangle ab is covered by the areas determined by the two i tegrals, but not all of the area of $\int_0^a f(x)dx$ is covered by tl rectangle.

Source: https://brilliant.org/wiki/youngs-inequality/



If $b > f(a)$ then the situation is similar, except that now not all of the area of $\int_0^b f^{-1}(y)dy$ is covered by the rectangle ab. Finally, if $b = f(a)$, the area covered by the two integrals matches the rectangle.

Step 4: We now apply the above to the function $y = f(x) = x^{p-1}$.

The inverse function is $x = f^{-1}(y) = y^{1/(p-1)} = y^{q-1}$ (see (11.42)). We integrate and obtain

$$\int_0^a f(x)dx = \int_0^a x^{p-1} = \frac{x^p}{p}\bigg|_0^a = \frac{a^p}{p}, \qquad \int_0^a f{-}1(y)dy = \int_0^a y^{q-1} = \frac{y^q}{q}\bigg|_0^b = \frac{b^q}{q}.$$

Young's inequality (11.41) now follows from (11.43). ■

Theorem 11.4 (Hoelder's inequality for L^p–norms)**.**

Let $a, b \in \mathbb{R}$ such that $a < b$. Let $p, q > 1$ be conjugate indices, i.e.,

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{11.44}$$

*Then **Hoelder's inequality** is true:*

(11.45)
$$\|fg\|_{L^1} \leq \|f\|_{L^p}\|g\|_{L^q}, \quad i.e., \int_a^b |f(x)g(x)|dx \leq \left(\int_a^b |f(x)|^p dx\right)^{1/p} \left(\int_a^b |g(x)|^q dx\right)^{1/q}.$$

PROOF: We note that the composite function $x \mapsto |f(x)|^p$ is continuous, hence integrable, as the composite of the three continuous functions $x \mapsto f(x), y \mapsto |y|$, and $z \mapsto z^p$.

Note that $\|f\|_{L^p} = 0$ is only possible if $|f(x)|^p = 0$, i.e., $f(x) = 0$ for all x (see prop.11.14 on p.530). Likewise, $\|g\|_{L^q} = 0$ implies $g(x) = 0$ for all x. In either case, $\int_a^b f(x)g(x) = 0$ and (11.45) is trivially satisfied. So we may assume that both $\|f\|_{L^p} > 0$ and $\|g\|_{L^q} > 0$ For some fixed $x \in [a, b]$ let

$$A := \|f\|_{L^p}, \ \ a_x := \frac{|f(x)|}{A}, \qquad B := \|g\|_{L^q}, \ \ b_x := \frac{|g(x)|}{B}.$$

It follows from Young's inequality (11.41) that $a_x b_x \leq \frac{{a_x}^p}{p} + \frac{{b_x}^q}{q}$. We integrate both sides of that inequality $\int_a^b \cdots dx$ and obtain from the monotonicity of the integral (see example ?? on p.??) that

$$\int_a^b a_x b_x dx \ \leq \ \int_a^b \Big(\frac{{a_x}^p}{p} + \frac{{b_x}^q}{q} \Big) \, dx. \tag{11.46}$$

i.e.,

$$\begin{aligned} \frac{1}{AB} \int_a^b |f(x)g(x)| \ dx \ &\leq \ \int_a^b \big(\frac{|f(x)|^p}{pA^p} + \frac{|g(x)|^q}{qB^q} \big) \ dx \\ &= \frac{1}{pA^p} \int_a^b |f(x)|^p \ dx \ + \ \frac{1}{qB^q} \int_a^b |g(x)|^q \ dx. \end{aligned} \tag{11.47}$$

We use

$$\int_a^b |f(x)|^p dx \ = \ (\|f\|_{L^p})^p, \quad \int_a^b |g(x)|^q dx \ = \ (\|g\|_{L^q})^q \tag{11.48}$$

in (11.47) and obtain

$$\frac{1}{AB} \int_a^b |f(x)g(x)| dx \ \leq \ \frac{A^p}{pA^p} + \frac{B^q}{qB^q} \ = \ \frac{1}{p} + \frac{1}{q} \ = \ 1.$$

It follows from the definition of A and B that

$$\int_a^b |f(x)g(x)| dx \ \leq \ AB \ = \ \|f\|_{L^p} \, \|g\|_{L^q}. \ \blacksquare$$

Theorem 11.5 (Minkowski's inequality for L^p–norms)**.**

Let $a, b \in \mathbb{R}$ such that $a < b$ and let $p \in [1, \infty[$. Then ***Minkowski's inequality*** *is true:*

$$\|f + g\|_{L^p} \ \leq \|f\|_{L^p} \ + \ \|g\|_{L^p}, \ \ \text{i.e.,} \tag{11.49}$$

$$\Big(\int_a^b |f(x) + g(x)|^p \ dx \Big)^{1/p} \ \leq \ \Big(\int_a^b |f(x)|^p \ dx \Big)^{1/p} \ + \ \Big(\int_a^b |g(x)|^p \ dx \Big)^{1/p}. \tag{11.50}$$

PROOF: This follows for $p = 1$ from part **(c)** of the proof of prop.11.17. We may assume that $p > 1$. Let q be the conjugate index to p, i.e.,

$$\frac{1}{p} + \frac{1}{q} \ = \ 1, \text{ hence } (p-1)q \ = \ p \tag{11.51}$$

(see (11.42)). Let $a \le x \le b$. Then

$$|f(x)+g(x)|^p = |f(x)+g(x)|\ |f(x)+g(x)|^{p-1} \le |f(x)|\ |f(x)+g(x)|^{p-1} + |g(x)|\ |f(x)+g(x)|^{p-1}.$$

The last inequality follows from $|f(x)+g(x)| \le |f(x)|+|g(x)|$ and $|f(x)+g(x)|^{p-1} \ge 0$. We integrate and obtain

$$\int_a^b |f(x)+g(x)|^p\,dx \le \int_a^b |f(x)|\ |f(x)+g(x)|^{p-1}\,dx + \int_a^b |g(x)|\ |f(x)+g(x)|^{p-1}\,dx. \tag{11.52}$$

We apply Hoelder's inequality to the first of the two integrals on the right–hand side of (11.52) and obtain

$$\begin{aligned} \int_a^b \big(|f(x)|\big)\ \big(|f(x)+g(x)|^{p-1}\big)\,dx &\le \Big(\int_a^b \big(|f(x)|\big)^p\,dx\Big)^{1/p} \Big(\int_a^b \big(|f(x)+g(x)|^{p-1}\big)^q\,dx\Big)^{1/q} \\ &= \Big(\int_a^b |f(x)|^p\,dx\Big)^{1/p} \Big(\int_a^b |f(x)+g(x)|^{(p-1)q}\,dx\Big)^{1/q} \\ &= \Big(\int_a^b |f(x)|^p\,dx\Big)^{1/p} \Big(\int_a^b |f(x)+g(x)|^p\,dx\Big)^{1/q}. \end{aligned} \tag{11.53}$$

The last equality results from $(p-1)q = p$ (see (11.51)). Similarly, we obtain from the second integral on the right–hand side of (11.51) the following:

$$\int_a^b \big(|g(x)|\big)\ \big(|f(x)+g(x)|^{p-1}\big)\,dx \le \Big(\int_a^b |g(x)|^p\,dx\Big)^{1/p} \Big(\int_a^b |f(x)+g(x)|^p\,dx\Big)^{1/q}. \tag{11.54}$$

We apply (11.53) and (11.54) to (11.52) and obtain

$$\begin{aligned} \int_a^b |f(x)+g(x)|^p\,dx &\le \Big(\int_a^b |f(x)|^p\,dx\Big)^{1/p} \Big(\int_a^b |f(x)+g(x)|^p\,dx\Big)^{1/q} \\ &+ \Big(\int_a^b |g(x)|^p\,dx\Big)^{1/p} \Big(\int_a^b |f(x)+g(x)|^p\,dx\Big)^{1/q}. \end{aligned} \tag{11.55}$$

Minkowski's inequality (11.50) is trivially satisfied if $\int_a^b |f(x)+g(x)|^p dx = 0$, so we may assume that $\int_a^b |f(x)+g(x)|^p dx > 0$. This allows us to divide each term in (11.55) by $\big(\int_a^b |f(x)+g(x)|^p dx\big)^{1/q}$. We obtain

$$\Big(\int_a^b |f(x)+g(x)|^p\,dx\Big)^{1-1/q} \le \Big(\int_a^b |f(x)|^p\,dx\Big)^{1/p} + \Big(\int_a^b |g(x)|^p\,dx\Big)^{1/p}. \tag{11.56}$$

Note that $1-\frac{1}{q} = \frac{1}{p}$ because $\frac{1}{q} + \frac{1}{p} = 1$, and (11.56) reads

$$\Big(\int_a^b |f(x)+g(x)|^p\,dx\Big)^{1/p} \le \Big(\int_a^b |f(x)|^p\,dx\Big)^{1/p} + \Big(\int_a^b |g(x)|^p\,dx\Big)^{1/p}. \ \blacksquare$$

Theorem 11.6 (Hoelder's inequality for the p–norms).

Let $n \in \mathbb{N}$ and $\vec{x} = (x_1, \dots x_N), \vec{y} = (y_1, \dots y_N) \in \mathbb{R}^n$. Let $p, q > 1$ be conjugate indices, i.e.,

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{11.57}$$

Then ***Hoelder's inequality*** *in $\mathbb{R}^n$ is true:*

$$\sum_{j=1}^n |x_j y_j| \leq \|\vec{x}\|_p \|\vec{y}\|_q, \quad i.e., \sum_{j=1}^n |x_j y_j| \leq \Big(\sum_{j=1}^n |x_j|^p\Big)^{1/p} \Big(\sum_{j=1}^n |y_j|^q\Big)^{1/q}. \tag{11.58}$$

PROOF: Let $\vec{x}, \vec{y} \in \mathbb{R}^n$. If $\vec{x} = 0$ or $\vec{y} = 0$ then $\sum_{j=1}^n |x_j y_j| = 0$ and (11.58) is trivially satisfied. We hence may assume that both $\vec{x} \neq 0$ and $\vec{y} \neq 0$.

It follows from part **(a)** of the proof of prop.11.16 (positive definiteness of $\|\cdot\|_p$ for all p) on p.533 that $\|\vec{x}\|_p > 0$ and $\|\vec{y}\|_q > 0$ For some fixed index $1 \leq j \leq n$ let

$$A := \|\vec{x}\|_p, \; a_j := \frac{|x_j|}{A}, \qquad B := \|\vec{y}\|_q, \; b_x := \frac{|y_j|}{B}.$$

It follows from Young's inequality (11.41) that

$$a_j b_j \leq \frac{a_j^p}{p} + \frac{b_j^q}{q}.$$

We take sums $\sum_{j=1}^n \cdots$ of both sides of that inequality and obtain from the monotonicity of summation

$$\sum_{j=1}^n a_j b_j \leq \sum_{j=1}^n \Big(\frac{(a_j)^p}{p} + \frac{(b_j)^q}{q}\Big), \tag{11.59}$$

i.e.,

$$\frac{1}{AB}\sum_{j=1}^n |x_j y_j| \leq \sum_{j=1}^n \Big(\frac{|a_j|^p}{pA^p} + \frac{|b_j|^q}{qB^q}\Big) = \frac{1}{pA^p}\sum_{j=1}^n |a_j|^p + \frac{1}{qB^q}\sum_{j=1}^n |b_j|^q. \tag{11.60}$$

But

$$\sum_{j=1}^n |x_j|^p = (\|f\|_p)^p, \quad \sum_{j=1}^n |y_j|^q = (\|g\|_q)^q. \tag{11.61}$$

It follows from (11.60) that

$$\frac{1}{AB}\sum_{j=1}^n |x_j y_j| \leq \frac{A^p}{pA^p} + \frac{B^q}{qB^q} = \frac{1}{p} + \frac{1}{q} = 1,$$

and we deduce from the definition of A and B that

$$\sum_{j=1}^n |x_j y_j| \leq AB = \|\vec{x}\|_p \, \|\vec{x}\|_q. \blacksquare$$

Theorem 11.7 (Minkowski's inequality for $(\mathbb{R}^n, \|\cdot\|_p)$)**.**

Let $n \in \mathbb{N}$ and $\vec{x} = (x_1, \dots x_N)$.
Let $\vec{y} = (y_1, \dots y_N) \in \mathbb{R}^n$ and $p \in [1, \infty[$. Then ***Minkowski's inequality for*** $(\mathbb{R}^n, \|\cdot\|_p)$ *is true:*

$$\|\vec{x} + \vec{y}\|_p \leq \|\vec{x}\|_p + \|\vec{y}\|_p, \text{ i.e.,} \tag{11.62}$$

$$\Big(\sum_j |x_j + y_j|^p\Big)^{1/p} \leq \Big(\sum_j |f(x)|^p\Big)^{1/p} + \Big(\sum_j |g(x)|^p\Big)^{1/p}. \tag{11.63}$$

PROOF: This follows for $p = 1$ from part **(c)** of the proof of prop.11.17. We hence may assume that $p > 1$. Let q be the conjugate index to p, i.e.,

$$\frac{1}{p} + \frac{1}{q} = 1, \text{ hence } (p-1)q = p \tag{11.64}$$

(see (11.42)). Let $a \leq x \leq b$. Then

$$|x_j + y_j|^p = |x_j + y_j|\ |x_j + y_j|^{p-1} \leq |x_j|\ |x_j + y_j|^{p-1} + |y_j|\ |x_j + y_j|^{p-1}.$$

The last inequality follows from $|x_j + y_j| \leq |x_j| + |y_j|$ and $|x_j + y_j|^{p-1} \geq 0$. We sum and obtain

$$\sum_j |x_j + y_j|^p \leq \sum_j |x_j|\ |x_j + y_j|^{p-1} + \sum_j |y_j|\ |x_j + y_j|^{p-1}. \tag{11.65}$$

Hoelder's inequality applied to the first of the two integrals on the right–hand side of (11.65) yields

$$\begin{aligned} \sum_j \big(|x_j|\big)\ \big(|x_j + y_j|^{p-1}\big) &\leq \Big(\sum_j \big(|x_j|\big)^p\Big)^{1/p} \Big(\sum_j \big(|x_j + y_j|^{p-1}\big)^q\Big)^{1/q} \\ &= \Big(\sum_j |x_j|^p\Big)^{1/p} \Big(\sum_j |x_j + y_j|^{(p-1)q}\Big)^{1/q} \\ &= \Big(\sum_j |x_j|^p\Big)^{1/p} \Big(\sum_j |x_j + y_j|^p\Big)^{1/q}. \end{aligned} \tag{11.66}$$

The last equality results from $(p-1)q = p$ (see (11.64)). Similarly, we obtain from the second integral on the right–hand side of (11.65) the following:

$$\sum_j \big(|y_j|\big)\ \big(|x_j + y_j|^{p-1}\big) \leq \Big(\sum_j |y_j|^p\Big)^{1/p} \Big(\sum_j |x_j + y_j|^p\Big)^{1/q}. \tag{11.67}$$

We apply (11.66) and (11.67) to (11.65) and obtain

$$\begin{aligned} \sum_j |x_j + y_j|^p &\leq \Big(\sum_j |x_j|^p\Big)^{1/p} \Big(\sum_j |x_j + y_j|^p\Big)^{1/q} \\ &\quad + \Big(\sum_j |y_j|^p\Big)^{1/p} \Big(\sum_j |x_j + y_j|^p\Big)^{1/q}. \end{aligned} \tag{11.68}$$

Minkowski's inequality (11.63) is trivially satisfied if $\sum_j |x_j + y_j|^p = 0$, so we may assume that $\sum_j |x_j + y_j|^p > 0$. We divide each term in (11.68) by $\left(\sum_j |x_j + y_j|^p\right)^{1/q}$ and obtain

$$\left(\sum_j |x_j + y_j|^p\right)^{1-1/q} \leq \left(\sum_j |x_j|^p\right)^{1/p} + \left(\sum_j |y_j|^p\right)^{1/p}. \tag{11.69}$$

Note that $1 - \frac{1}{q} = \frac{1}{p}$ because $\frac{1}{q} + \frac{1}{p} = 1$, and (11.69) reads

$$\left(\sum_j |x_j + y_j|^p\right)^{1/p} \leq \left(\sum_j |x_j|^p\right)^{1/p} + \left(\sum_j |y_j|^p\right)^{1/p}. \blacksquare$$

In chapter **??** (Metric Spaces and Topological Spaces – Part I) you will learn about metric spaces as a concept that generalizes the measurement of distance (or closeness, if you prefer) for the elements of a nonempty set.

11.3 Exercises for Ch.11

Exercise 11.1.

Prove prop.11.13 on p.526: Let X be an arbitrary, nonempty set. Then the function $\|\cdot\|_\infty : \mathscr{B}(X,\mathbb{R}) \to \mathbb{R}_+$, $\quad h \to \|h\|_\infty = \sup\{|h(x)| : x \in X\}$ defines a norm. $\square$

Exercise 11.2.

Prove parts **(a)** and **(b)** of prop.11.10 (Algebraic properties of the inner product) on p.524:

Let V be a vector space with inner product $\bullet(\cdot,\cdot)$. Let $a, b, x, y \in V$. Then

- **(a)** $(a+b) \bullet (x+y) = a \bullet x + b \bullet x + a \bullet y + b \bullet y$
- **(b)** $(x+y) \bullet (x+y) = x \bullet x + 2(x \bullet y) + y \bullet y$
- **(c)** $(x-y) \bullet (x-y) = x \bullet x - 2(x \bullet y) + y \bullet y$ $\square$

Exercise 11.3. Prove Remark 11.8 on p.527:

Let $(V, \|\cdot\|)$ be a normed vector space and let $\gamma > 0$. Let $p : V \to \mathbb{R}$ be defined as $p(x) := \gamma\|x\|$. Then p also is a norm. $\square$

Exercise 11.4.

Prove that the p–norm (see Definition 11.18 on p.527) is a norm on $\mathbb{R}^n$ for the special case $p = 1$:

$$\|\vec{x}\|_1 = \sum_{j=1}^{n} |x_j| \quad \square$$

References

[1] Matthew Brin and Gerald Marchesi. Linear Algebra, a text for Math 304, Spring 2016. Binghamton University, 13th edition, 2016.

[2] A.N. Kolmogorov and S.V. Fomin. Introductory Real Analysis. Dover, Mineola, 1st edition, 1975.

[3] James Stewart. Single Variable Calculus. Thomson Brooks Cole, 7th edition, 2012.

[4] Richard E. Williamson, Richard H. Crowell, and Hale F. Trotter. Calculus of Vector Functions. Prentice Hall, Englewood Cliffs, 3rd edition, 1972.

List of Symbols

Index